Journal of Memory and Language 93 (2017) 1–21

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Exploring the relations between word frequency, language exposure, and bilingualism in a computational model of reading

Padraic Monaghan [a,b,*], Ya-Ning Chang [a], Stephen Welbourne [c], Marc Brysbaert [d]

[a] Lancaster University, UK
[b] Max Planck Institute for Psycholinguistics, Netherlands
[c] University of Manchester, UK
[d] Ghent University, Belgium

ARTICLE INFO

Article history:
Received 14 December 2015
revision received 24 August 2016
Available online 12 September 2016

Keywords:
Reading
Frequency effects
Computational modelling
Individual differences
Bilingualism
Lifespan development

ABSTRACT

Individuals show differences in the extent to which psycholinguistic variables predict their responses for lexical processing tasks. A key variable accounting for much variance in lexical processing is frequency, but the size of the frequency effect has been demonstrated to reduce as a consequence of the individual's vocabulary size. Using a connectionist computational implementation of the triangle model on a large set of English words, where orthographic, phonological, and semantic representations interact during processing, we show that the model demonstrates a reduced frequency effect as a consequence of amount of exposure to the language, a variable that was also a cause of greater vocabulary size in the model. The model was also trained to learn a second language, Dutch, and replicated behavioural observations that increased proficiency in a second language resulted in reduced frequency effects for that language but increased frequency effects in the first language. The model provides a first step to demonstrating causal relations between psycholinguistic variables in a model of individual differences in lexical processing, and the effect of bilingualism on interacting variables within the language processing system.

Introduction

Word frequency is a key variable in predicting differences in word processing efficiency: High frequency words are recognized faster and more accurately than low frequency words (Forster & Chambers, 1973). Measured against a range of other psycholinguistic properties, frequency accounts for a far larger amount of variance in response times and accuracies than other variables. For instance, in one of the earlier "mega-studies" of word processing, Balota, Cortese, Sergent-Marshall, Spieler, and Yap (2004) found that frequency exceeded neighbourhood size and consistency in explaining variance of response times for word naming, and matched the size of the effect of word length. For lexical decision, they found that the standardized regression coefficient for frequency was at least four times as great as that of any other psycholinguistic variable (for other regression analyses demonstrating a similarly greater effect of frequency, see Brysbaert, Stevens, Mandera, & Keuleers, 2016; Brysbaert et al., 2011; Cortese & Khanna, 2007; Keuleers, Stevens, Mandera, & Brysbaert, 2015; Spieler & Balota, 1997; Yap & Balota, 2009). Frequency is taken to indicate greater efficiency of access, more salient representation of the lexical item, and greater availability of the representation within the individual's vocabulary (Adelman, Brown, & Quesada, 2006).

* Corresponding author at: Department of Psychology, Lancaster University, Lancaster LA1 4YF, UK. Fax: +44 1524 593744.

E-mail address: p.monaghan@lancaster.ac.uk (P. Monaghan).

http://dx.doi.org/10.1016/j.jml.2016.08.003
0749-596X/

The frequency effect is typically treated in analyses as a random effect as if variance across participants is random. Hence, until very recently, frequency effects have tended to have been related to mean group responses to individual words, rather than appraised in terms of individuals responding to individual words. However, in the first study on the phenomenon it was already reported that the frequency effect differed between participants who had small and large vocabularies. In a largely overlooked paper, Preston (1935) was the first to examine the word frequency effect. She measured the 'speed of word perception' for familiar and unfamiliar words of the same length. The stimulus words consisted of 50 familiar and 50 unfamiliar six-letter two-syllable words chosen on the basis of Thorndike's (1931) 20,000 Word List. The familiar words were selected from the 1500 highest words of the list (i.e., those used most frequently in printed matter). The unfamiliar words were selected from the 19th and the 20th thousand lowest words. Speed of word perception was "measured by the time between the exposing of a stimulus word and the verbal reading of it" (nowadays called a word naming task). Eighty-one members of elementary psychology classes at the University of Minnesota served as participants. Their average "perception time" for the familiar words was 578 ms; that for the unfamiliar words 691 ms.

A second purpose of Preston's study was "the study of the relation of various measures of reading ability to speed of word perception."[1] The reading ability of the participants was determined by the administration of the Vocabulary Test of the Minnesota Reading Examination, the Chapman Cook Speed of Reading Test, and Test V of the Iowa Silent Reading tests. The first test contained 100 words with five possible definitions from which examinees had to select the correct definition. In the Chapman Cook Speed of Reading Test participants were presented with 25 short paragraphs in which one word spoiled (sic) the paragraphs. Participants had to find as many intruder words as possible in 2.5 min and cross out these words. Test V of the Iowa Silent Reading tests was a paragraph comprehension test, in which 12 paragraphs had to be read and 3 questions answered per paragraph. Preston observed significant negative correlations between the language proficiency test scores and the word perception response times, with the highest correlation between vocabulary size and word perception response times, and the lowest correlation between text comprehension and word perception response times. The correlation was higher for the unfamiliar words than the familiar words (e.g., the correlation between vocabulary size and word perception response time was −.508 for the unfamiliar words, and −.412 for the familiar words). In other words, the relation between vocabulary size and response times was greater for low- than high-frequency words, suggesting that individual differences in reading responses may reduce as a consequence of exposure.

Preston's (1935) paper was not mentioned in Howes and Solomon's (1951) article examining the relationship between word frequency and visual duration thresholds in a word identification task. This publication is (erroneously) considered to be the start of word frequency research by many researchers. In two experiments, Howes and Solomon presented evidence that the visual duration threshold in word identification decreased as a function of the logarithm of word frequency (also based on Thorndike's counts). Importantly, and unfortunately, no individual differences were examined and the word frequency effect was presented as a group effect, assumed to be observed to the same degree in all participants. Howes and Solomon's view has dominated the literature, even though occasionally differences in the frequency effect between groups have been investigated (e.g., Chateau & Jared, 2000; Lewellen, Goldinger, Pisoni, & Greene, 1993; Sears, Siakaluk, Chow, & Buchanan, 2008).

Our own interest in individual differences in the word frequency effect arose from a series of experiments published by Yap, Balota, Tse, and Besner (2008).[2] In this article the authors presented data from three different universities on the same lexical decision task. Table 1 gives a summary of the finding that caught our attention. As in Preston's (1935) study, students with a smaller vocabulary size had longer reaction times and, more importantly, showed a larger frequency effect.

The influence of vocabulary size on the frequency effect was later replicated in a large-scale analysis of individual differences in the English Lexicon Project (Yap, Balota, Sibley, & Ratcliff, 2012).

At first sight, it seems surprising that people with a larger vocabulary are more efficient at activating the correct representation than those with a smaller vocabulary, given that they have to select among more candidates in the vocabulary (Lewellen et al., 1993). Still, there are at least four mechanisms that may contribute to the effect. The first is that a larger frequency effect may be a side-effect of longer reaction times (RTs; Faust, Balota, Spieler, & Ferraro, 1999): Comparing the data from Yap et al. (2008) shown in Table 1, 678 ms is 11% longer than 612, and 844 is 15% longer than 732 ms. If we assume that part of the RT to words is not due to word processing but to constant durations such as those involved in stimulus transmission and action planning and performance, it could even be possible that the proportional increase between low and high frequency words is the same across the groups. For the example at hand, this would be the case when the constant time period for stimulus transmission and action is around 438 ms, as then for the lowest vocabulary group the stimulus processing time would be 240 ms [678–438], and 174 ms for the highest vocabulary group, which is 38% different. For the high frequency words, the differences between the highest and lowest vocabulary group would be 406 ms and 294 ms, which is again 38% more. Thus, it is feasible that vocabulary size affects word processing speed generally, rather than affecting the variance associated with the frequency effect.

[1] There was also a third purpose: To determine the test-retest reliability of the speed of word perception measure by asking participants to name the words twice with six days or more in-between. The reliability was .93.

[2] Just like many other researchers, we were until recently unaware of the Preston (1935) paper. We thank Andy Ellis for pointing it out to us.

Table 1
Frequency effect of 3 groups of students with different vocabulary sizes on the same lexical decision task, based on Yap et al. (Experiments 2–4, clear presentation condition).

University	Vocabulary[a]	RT_{LF} (ms)	RT_{HF} (ms)	Effect (ms)
Washington U.	18.7	678	612	66
Waterloo	17.7	753	658	95
Albany (SUNY)	16.9	844	732	112

[a] As determined with the Shipley (1940) vocabulary test: Vocabulary age is estimated on the basis of 40 words with 4 response alternatives each.

A second explanation for individual differences in the frequency effect could be that the more efficient retrieval operation in people with large vocabulary sizes is due to their higher intelligence. Indeed, vocabulary tests are used as a part of measures of intelligence, and load on g (Wechsler, 2008), and g in turn relates to processing speed (Salthouse, 1996). So, the relation between the frequency effect and vocabulary size could be an artefact of intelligence. However, this interpretation received a serious setback when it was observed that exactly the same function accounts for the relation between vocabulary size and frequency effects in second language (L2) processing as in first language (L1) processing (Brysbaert, Lagrou, & Stevens, in press; Diependaele, Lemhöfer, & Brysbaert, 2013). The frequency effect is larger in L2 than L1, but this difference disappears when vocabulary size is taken into account. The apparently larger effect of frequency in L2 is thus because people generally know fewer words in L2 than in L1. It is difficult to maintain that people would be less intelligent in L2.

A third possible contribution to the correlation between vocabulary size and the frequency effect relates to differences in the type of input. Some people may be exposed to more varied input than others. For instance, it is well established that written language comprises a more varied vocabulary than spoken language (for reviews, see Kuperman & Van Dyke, 2013; Pfost, Dörfler, & Artelt, 2013), at least partially because word repetition is tolerated in speech but not in writing. However, even when modality of input is controlled, Kuperman and Van Dyke (2013) showed that a larger input is associated with relatively more exposure to low frequency words.

Finally, it could be the case that higher exposure by itself is enough to explain the smaller word frequency effect, without any need for extra variables. In that scenario, both the small frequency effect and the large vocabulary size would be consequences of language exposure, which has a larger effect on the efficiency of word retrieval than on the cost of interword competition. Such a view would be by far the simplest interpretation and, hence, it is worthwhile to examine whether it can be observed in computational models of word processing.

The type of computational model best suited to investigate learning effects consists of the distributed connectionist models (Chang, Furber, & Welbourne, 2012; Harm & Seidenberg, 2004; Monaghan & Ellis, 2010; Plaut, McClelland, Seidenberg, & Patterson, 1996; Welbourne & Lambon Ralph, 2007). In these models, words are not represented as localist representations (nodes in a network), but as activation patterns across orthographic, phonological and semantic layers. The connection weights between the layers determine the efficiency with which one representation can activate the other. These depend on a number of factors, including the number of times an item has been presented to the model. Stimuli that are often presented succeed in a greater accumulation of adaptation of the weights in the network, so that the output they generate resembles the desired output to a closer extent. In contrast, stimuli with a low presentation probability have less impact on the organisation of the network and take more time to be effectively learned, resulting in larger error as the model attempts to produce phonological or semantic representations from a given orthographic input. As a result, distributed networks are able to simulate frequency effects without any requirement of the researcher to introduce a frequency dependent parameter (see, e.g., Harm & Seidenberg, 1999; Seidenberg & McClelland, 1989). In these models, high frequency words are processed more accurately than low frequency words because the connections supporting learning the mapping between orthographic, phonological, and semantic representations have undergone more adjustment to reduce error within the system for the higher frequency words. Thus, the model processes words to which it has been exposed with greater fidelity. Accuracy of production of phonological (for word naming) or semantic (for lexical decision) representations has been taken to reflect response times in behavioural lexical processing in previous models (Plaut et al., 1996; Seidenberg & McClelland, 1989).

The triangle model refers to the connectionist model where orthographic, phonological, and semantic representations interact in word processing (Harm & Seidenberg, 2004; Seidenberg & McClelland, 1989). This model has been tested on a range of group level effects, such as word frequency, yet it also has the potential to reflect individual differences in performance. In particular, the various theories about the relation between vocabulary knowledge, first and second language facility, and exposure can be tested for the extent to which they give rise to frequency effects within the model.

There are alternative models that could also potentially be used to test these individual differences in performance. The dual route cascaded (DRC) model implements two routes for mapping from orthography to phonology, a sublexical route that maps letters to sounds via a set of grapheme-phoneme correspondence rules, and a lexical route containing word units which directly, and simultaneously, activate the phonology corresponding to the whole word (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). Such models implement word frequency effects by adding an inhibitory bias that is inversely proportional to the log of the frequency of the word. Adelman and Brown (2008) showed how variables within this model could be systematically varied to test fit of the model to data, and Ziegler et al. (2008) tested the extent to which adjusting variables in the DRC model could simulate individual variation in reading as a consequence of visual letter, word-level phonological, and segmental phonological skills. In the case of the frequency effect, this could be adjusted within the DRC model by varying the gradient of the frequency bias inhibitory function, or by varying the frequencies of words in the model's input, as a proxy to adjusting the model's environment. A third alternative would be to vary the relative contribution of the lexical and sublexical routes to word naming performance. As the sublexical route is not affected by individual word frequency, word frequency effects would be reduced if the sublexical route contributes more to the model's response. However, these effects would have to be implemented in the model, rather than be an emergent consequence of the way the reading system interacts with the environment. A more recent instantiation of a dual route model, comprising lexical and sublexical routes, is the CDP+ model (Perry, Ziegler, & Zorzi, 2007). For this model, the lexical route is similar to that of the DRC, but the sublexical route learns to adjust weights between particular letters and phonemes according to their relative frequencies. Consequently, frequency effects at the word level are again implemented within the lexical route, but the overall size of the frequency effect could again be altered by varying the relative contribution of lexical and sublexical routes to performance. In Adelman, Sabatos-DeVito, Marquis, and Estes' (2014) test of individual differences within the CDP+ framework, they interpreted frequency effects as emerging only from the former variable: via adjustment of the frequency inhibitory bias in the lexical route.

Our aim in this paper is to determine the extent to which quantitative changes in exposure to words can affect the frequency effect in word naming. We report the results from a series of simulations systematically examining the size of frequency effects during training of the connectionist triangle model of reading (Harm & Seidenberg, 2004; Seidenberg & McClelland, 1989). Examining the triangle model enables us to ascertain the extent to which exposure alone has an effect on frequency effects, without imposing adaptations to the system, as would be the case using the DRC or CDP+ models as starting points. The precise characteristics of the model we view as not being the critical issue, but rather we provide an exploration of the principle of how environment can impact on psycholinguistic factors affecting word representation.

In Simulation 1, we determined whether the behavioural observation of the reduced frequency effect relating to vocabulary size may be a consequence of greater exposure to the vocabulary in the model. We tested whether exposure results in decreasing frequency effects for both naming (simulated by orthography to phonology mappings within the model) and lexical decision (simulated by orthography to semantics mappings). We anticipate that reductions in the frequency effect may result from increasing the efficiency of mappings in the model, as a consequence of extended exposure to the vocabulary. We further tested whether the model's performance is due to a linear improvement in responding to all words, or whether a reduced frequency effect may be caused by improved fidelity of low frequency word mappings.

Simulations 2 and 3 teased apart the relative contribution of vocabulary exposure and vocabulary size, by training the model on different vocabulary sizes. We predicted that vocabulary exposure would be the key factor resulting in changes in the frequency effect. Finally, Simulation 4 tested the effect of learning a second language on frequency effects in the model, and whether increasing proficiency in the second language resulted in reduced frequency effects in this second language and increased frequency effects in the first language, as a result of vocabulary size differences, in turn resulting from differences in exposure to the two languages. For this simulation, we introduced a second language – Dutch – to the triangle model in order to investigate the relative frequency effects within the model for its reading of English and Dutch words, as exposure to each language varied.

Simulation 1: frequency effects in the triangle model of reading

Method

Architecture

The model was based on the connectionist triangle model of Harm and Seidenberg (2004), and is shown in Fig. 1. The model comprised three representational layers, where orthographic, phonological, and semantic representations of words were presented. It was limited to monosyllabic words.

The phonological layer was connected to and from a set of 50 cleanup units to enable the model to develop stable phonological representations for words. The phonological layer was connected to the semantic layer via a set of 300 hidden units. The semantic layer was connected to and from a set of 50 semantic cleanup units. The semantic layer was connected to the phonological layer via another set of 300 hidden units.

A 4 unit context layer was connected to the semantic layer via a set of 10 hidden units. This context layer enabled the model to disambiguate homophones using context. For each homophone, a different context unit was active. Which unit was active for each set of homophones was selected randomly, such that each context unit was active to approximately the same frequency across the training set. For words which were not homophones, all context layer units were inactive.

The orthographic layer was connected to the phonological layer via a set of 100 hidden units, and to the semantic layer via a set of 300 units. A different number of units was required for successfully learning the mapping from orthography to phonology than for orthography to semantics (see Plaut et al., 1996, for requirements of learning pseudo-regular and arbitrary mappings).

Training set

Written forms of monosyllabic words were presented at the orthographic layer, which comprised 10 letter slots,

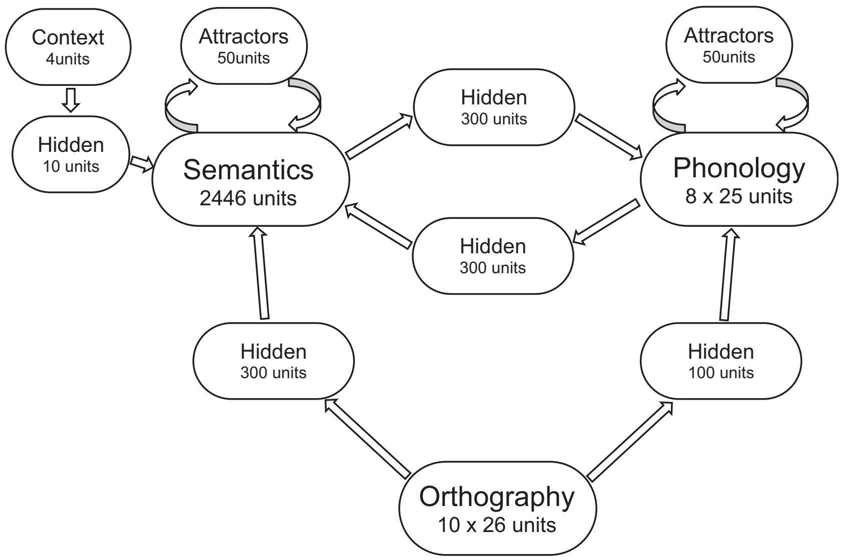

Fig. 1. Architecture of the triangle model of reading used in the current simulations.

within which each letter was represented as one unit active from a set of 26. Words were vowel-centred, such that the first vowel in the word was presented at the fourth letter slot, with two slots available for up to two consecutive vowels in the orthography. Consonants preceding the vowel were presented across slots 1–3, with these onset consonants in adjacent slots to the vowel. The remaining consonants and following vowels were presented in slots commencing at slot 7 and filled slots adjacent to the two vowel slots. Thus, for the word "plane", the orthographic representation was presented across the slots _ p l a _ n e _ _ _, and for "aunt", the orthographic representation was _ _ _ a u n t _ _ _. A letter present in each position was represented as the unit in the slot associated with that letter having activity 1.

Phonological forms of words were presented at the phonological layer, which comprised 8 phoneme slots, with each slot composed of a set of 25 phonological features. Phonological features were exactly those used by Harm and Seidenberg (2004). Phonological representations of words were presented with three slots for the onset, one slot for the vowel, and four slots for the coda. Onset and coda consonants were presented across slots directly adjacent to the vowel. Diphthongs, and long and short vowels were all represented as a set of features active in a single vowel slot. So, for the word "plane", the phonological representation was _ p l eI n _ _ _. Phoneme features had activity 1 in phoneme slots that were present in the input.

The semantic representations of each word were acquired from Wordnet (Miller, 1990), using the same algorithm described by Harm and Seidenberg (2004). The semantic representation for each word comprised an activated subset of 2446 semantic features. Presence of a feature was represented with activity 1.

There were a total of 6229 words, which comprised all monosyllabic words in English which had both a phonological representation in the CMU pronouncing dictionary (Weide, 1998) and a semantic representation listed in Wordnet (Miller, 1990). This set of words was slightly greater than that used in Harm and Seidenberg (2004) because in their simulations they only included word forms with their most frequent inflected form, whereas we included all monosyllabic inflected versions of the word.

Frequency of words was derived from the Wall Street Journal corpus (Marcus, Santorini, & Marcinkiewicz, 1993), and frequency was log-compressed prior to training of the model. This measure of frequency was that employed in the first implementation of the triangle model (Harm & Seidenberg, 2004), and is included here for comparison with this earlier version. Note that this compression maintains the relative frequency order of words, but substantially reduces the range of frequencies for the model. The model therefore applies a stringent test of the extent to which the changing frequency effects in behaviour can be simulated with this smaller distinction between word frequencies.

Training and testing

Five versions of the model were trained as separate simulations, with different randomised starting weights, and different random orderings of training patterns selected according to frequency. This ensured that the observed results were not due to particular starting configurations of the model.

Pretraining. The model was first trained to learn to map between phonological and semantic representations, as well as to develop stable phonological to phonological mappings, and semantic to semantic mappings.

For the phonological to phonological mapping trials, a phonological representation of a word was presented at the phonological layer. Then, the activity in the model was allowed to cycle for 6 time steps, and for time steps

7 and 8 the model was required to reproduce the phonological representation of the word. Similarly, for the semantic to semantic trials, the model was required to reproduce the semantic representation presented at the semantic layer in time steps 7 and 8. For the phonological to semantic mappings, the phonological representation and the context representation was presented to the model for a word for all 8 time steps, and the model was required to produce the semantic representation of the word in time steps 7 and 8. For semantics to phonological mappings, the semantic representation was presented at all time steps, and the model was required to produce the phonological representation of the word at time steps 7 and 8. As the semantic representation was unambiguous with respect to producing the phonological form of the word, the context layer was not necessary in order to form this mapping.

The model was trained using recurrent backpropagation, with cross-entropy error computed between the target and the model's actual production for each word's representation. The learning rate was set at 0.05. The pretraining comprised 2 million word presentations, with words selected according to their log-compressed frequency, in the range [0.05, 1]. 10% of trials were the phonological to phonological mapping, 10% were semantics to semantics, 40% of trials mapped from semantics to phonology, and the remaining 40% mapped from phonology to semantics.

Reading training. Following pretraining, the model then learned to map from orthographic forms onto phonological and semantic representations. The orthographic representation of a word was presented at the orthographic layer, and simultaneously the context layer representation was also presented. Then, from time steps 7 to 12, the model was required to produce the phonological and the semantic representation for that word. Cross-entropy error was backpropagated through the model, and the learning rate was set at 0.01. The model was trained for 1 million presentations.

Testing. The pretraining model was tested on both phonological to semantic trials, and semantic to phonological trials. For the phonological to semantic trials, the phonological representation of each word was presented, and then the model's production at the semantic layer at the end of the 8 time steps of activation was recorded. The closeness of the model's semantic production was determined by measuring the sum squared error over the semantic layer. The accuracy of the model's semantic production was measured by computing the cosine of the model's actual semantic representation against the semantic representations of each of the 6229 words in the training set. If the cosine distance was lowest for the target representation then the model was judged to be accurate.

For the semantic to phonological trials, the semantic representation was presented and then the phonological production was compared to the target phonological representation after 8 time steps, then the closeness of the model's production was determined by measuring sum squared error. Accuracy of the model was measured by determining for each phoneme slot the closest phoneme to the model's actual production. If the closest phonemes matched the target in all positions then the model's phonological production was judged to be accurate.

For the reading trials, the model was presented with the orthographic representation of each word, and closeness and accuracy of the model's actual production at both the semantic and the phonological layers were recorded. As with behavioural studies of reading, we distinguish accuracy of responses from response time measures. The model may produce an accurate response (closer to the target than any other representation in the training set) but to varying degrees of closeness in terms of the actual versus target representation. Closeness of the model's phonological production to the target phonology was taken to relate to response time measures of naming, in accordance with previous connectionist models of reading (e.g., Harm & Seidenberg, 2004; Monaghan & Ellis, 2010; Plaut et al., 1996) as it provides an indication of the ease with which the model can generate the phonological form of the word from its orthographic input. Similarly, the closeness of the semantic production was related to response times in lexical tasks involving generation of a semantic representation, as again the closeness reflects the ease with which the model can produce a meaning representation from orthographic input.

An alternative measure of accuracy of semantics would be to determine whether each feature was activated above or below a given threshold, rather than to measure accuracy based on relative distance to other patterns in the training set. To determine whether taking a unit threshold of 0.5 at the semantic output layer resulted in a different reflection of accuracy, we compared the model's performance for the nearest neighbour and threshold function accuracy measures. At the end of training, the model was able to solve the task to a high degree of accuracy for both accuracy measures (for nearest neighbour: mean = 99.7%, *SD* = .05%, for threshold: mean = 98.3%, *SD* = .07%). There was a high degree of correspondence between the threshold measure of accuracy and the nearest semantic representation measure: mean agreement = 98.5% of patterns, *SD* = .6%, $\chi^2(1) = 3828.3$, $p < .0000001$. Thus, the model was able to solve the mapping task to a high degree of accuracy regardless of the precise measure of accuracy.

Results

The model's performance for accuracy was assessed using generalized linear mixed effects models, and measures for frequency effects were assessed on the model's error. The significance of individual and interacting factors was assessed by determining whether the model fit improved significantly by applying a likelihood ratio test comparison between models with and without the factor or interaction of interest.

Pretraining

Pretraining was halted after 2 million patterns, and at this point the model achieved mean accuracy of 96.0% (*SD* = 1.9%) for mapping from semantic to phonological representations, and 87.8% (*SD* = 1.2%) for mapping from phonological to semantic representations (see Fig. 2). To

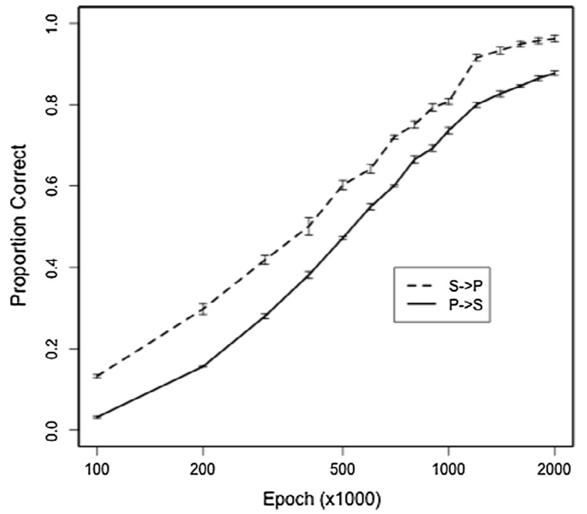

Fig. 2. Performance of the triangle model during pretraining between phonological and semantic representations (S → P is semantics to phonology mappings, P → S is phonology to semantics mappings). Error bars show ±1 SEM of mean accuracy by simulation.

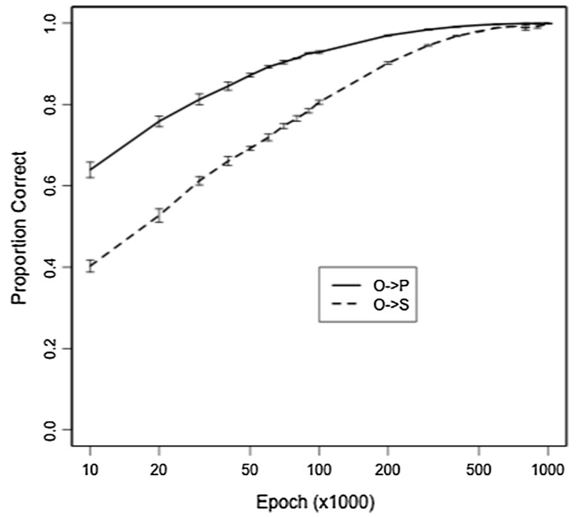

Fig. 3. Performance of the triangle model during training on orthography to phonological (O → P) and orthography to semantic (O → S) representations. Error bars show ±1 SEM of mean accuracy by simulation.

test whether semantic representations were slower to acquire than phonological representations during learning, we compared the fit of binary logistic linear mixed effects models. As a baseline, we constructed a model with simulation (simulation one to five) and word (each of the 6229 vocabulary items) as random effects, and log of training epoch as a fixed effect, with accuracy (correct or incorrect) of the model as the dependent variable. We then tested whether adding mapping type (semantics to phonology, or phonology to semantics) to this model resulted in a significant improvement of fit. We found that it did, $\chi^2(1) = 28{,}851$, $p < .001$, thus, the computational model learned to map from semantics to phonology more accurately than phonology to semantics. This was likely because the semantic input representations were more distinct, enabling greater differentiation of input patterns during training.

Reading accuracy

For the full reading model, accuracy for mapping from orthography to phonology and to semantics is shown in Fig. 3. By the end of 1 million patterns of training, the model was able to accurately produce the phonological (mean = 99.9%, *SD* = .03%) and the semantic representations (mean = 99.8%, *SD* = .05%). A binary logistic mixed effects model with simulation and word as random effects, and log of training epoch as fixed effect was improved in fit by adding in an additional fixed effect of mapping type (orthography to phonology, or orthography to semantics), $\chi^2(1) = 47{,}542$, $p < .001$. Adding an interaction between training epoch and mapping type also improved fit significantly, $\chi^2(1) = 244.24$, $p < .001$, indicating that phonological representations were learned more accurately than semantic representations especially in the early stages of training.

Frequency effects

To determine the extent to which frequency effects varied as a consequence of exposure, the correlation between frequency and the closeness of the model's output production compared to the target representation, as measured by mean square error, for phonological and semantic representations is shown in Fig. 4. Frequency effects can then be determined by the extent to which the frequency of a word improves the fit of the statistical model to the computational model error data. Changes in the frequency effect can then be determined by examining the interaction of frequency with other fixed factors in the model.

To compare frequency effects across the phonological and semantic representations, a mixed effects model with simulation and word as random effects, and log of training epoch as fixed effect was constructed as a baseline. Adding mapping (orthography to phonology, or orthography to semantics) as a fixed effect improved model fit, $\chi^2(1) = 246{,}635$, $p < .001$, as did adding word frequency, $\chi^2(1) = 1920.6$, $p < .001$. This indicated that, overall, there was a frequency effect in the triangle model's performance. Adding the interaction between frequency and mapping also improved fit, $\chi^2(1) = 70{,}012$, $p < .001$. This indicated that, as anticipated, the frequency effect was larger for the semantic representations than for the phonological representations. This is consistent with a greater effect of item-level properties for arbitrary than for consistent mappings, both within mappings, such as in the frequency by consistency effect for single word naming tasks (Taraban & McClelland, 1987) and across mappings, such as the larger frequency effect as a predictor of lexical decision response times (which has been proposed to involve semantic representations) compared to naming times for single words (Ghyselinck, Lewis, & Brysbaert, 2004).

In general, the frequency effect for both semantic and phonological representations declined with length of training. For instance, for the semantic representations change

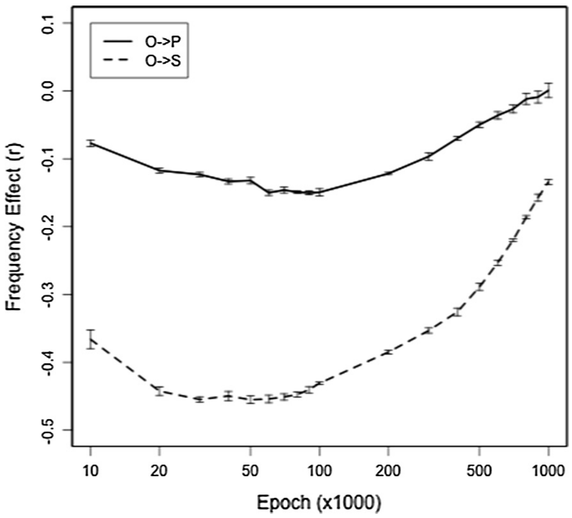

Fig. 4. Frequency effect for orthographic to phonological (O → P) and orthographic to semantic (O → S) representations across learning. The frequency effect initially increases in magnitude (i.e., the negative correlation between frequency and output quality becomes stronger) as a function of practice and then decreases to a lower level. The frequency effect is larger for orthography to semantics than for orthography to phonology. Error bars show ±1 SEM of mean correlation between word frequency and error by simulation.

in correlation from 100,000 training patterns (mean $r = -.431$, averaged over the five simulations) to 1 million reading patterns (mean $r = -.134$) was significant, $Z = 18.19$, $p < .001$. Similarly, for the phonological representations, change from 100,000 (mean $r = -.149$) to 1 million (mean $r = .008$) was significant, $Z = 8.87$, $p < .001$. Thus, in the triangle model the reduction in frequency effect was consistent with the theoretical proposal that exposure to words results in a reduction of the frequency effect.

However, the decline in the frequency effect was not monotonic. Instead the effect of frequency with training seemed to demonstrate a U-shaped curve. Over the early stages of training, the frequency effect gradually increased in magnitude before decreasing from approximately 100,000 training patterns onwards. To test whether a quadratic curve was a better fit to the data than a linear function, we compared two models for the phonological and the semantic representations separately. The first model was a linear mixed effects model with an interaction between a linear effect of log epoch of training and frequency, and the second was a linear mixed effects model with an interaction between a quadratic effect of log epoch of training and frequency. Simulation and word were random effects. For the phonological representations, the quadratic model was a better fit than the linear model, $\chi^2(2) = 17{,}599$, $p < .001$. The quadratic model was also a better fit for the semantic mappings, $\chi^2(2) = 11{,}532$, $p < .001$.

It is possible, then, that the frequency effect has two components: First, an effect of noise reduction related to learning to generate stable and accurate representations within phonology and within semantics, which occurs during the early stages of learning. Initial increases in fidelity result in increasing frequency effects. Then, second, continued exposure to the stimuli results in a gradual erosion of the frequency effect. Such an effect would be hard to explain with a single component view of the effect of frequency, in which case we would expect a better fit of a linear relation between frequency and training time on accuracy levels. Fig. 5 shows the model's error in producing phonological and semantic representations for every word in the corpus, averaged over the five simulations at different stages of training. At 10,000 epochs of training, the model demonstrated high error rates for patterns at all frequencies. After 100,000 epochs of training, error is predominantly for lower word frequencies. After 1 million epochs of training, the variation for low frequency words has also reduced, indicating that the reduction in the frequency effect is due to reduced error for the low frequency items as a result of exposure (see Diependaele et al., 2013, for similar patterns in behavioural responses, but see Kuperman & Van Dyke, 2013, for an alternative perspective). Fig. 6 summarises these data by showing how the frequency effect changes with training time for higher (frequency >0.5) and lower (frequency ⩽.5) frequency words.

The reduction in frequency effects for later training is therefore consistent with Plaut et al.'s (1996) account of how accuracy of reproduction of an individual word relates to training exposure in associative learning models, and consequently reduces the opportunity for other variables to contribute variance to the model's performance for mapping a particular pattern. This is because, as the model gets closer to producing the target representation at the output the slope of change of the activation function gets shallower, meaning that the contribution of competing, or interfering, factors that are general to the set of patterns learned, rather than the particular pattern being processed, have less of an impact on performance. Such an effect is a result of error-driven learning mechanisms, such that, as error reduces for the higher-frequency words, adjustments in the model are driven more and more by the lower-frequency words which are producing comparatively higher error rates. As reflected in Fig. 6, early in training the model's weight changes are driven more by the higher frequency words as these occur more often in the model's exposure, but then as error rates reduce for mapping the higher frequency words, the lower frequency words then begin to drive error rates, resulting in a later decrease in the frequency effect for these lower-frequency words.

The consequence of this error-driven learning is that, as the model produces representations closer to the target representation, the variation in error also declines. Thus, the reduction in the frequency effect is akin to a "floor" effect in performance. For instance, across all five simulations, at 10,000 epochs of training for the model's phonological output, mean error = 1.23, variance = 3.35. By 100,000 epochs, mean = .184, variance = .506, and by 1,000,000 epochs, mean = .00179, variance = .00371. Whereas the ratio of mean to variance remains similar, the size of the mean error reduces greatly. Our model is a simulation of the processes involved in lexical access, whereas other models of lexical processing may also consider the decision making processes involved in generating

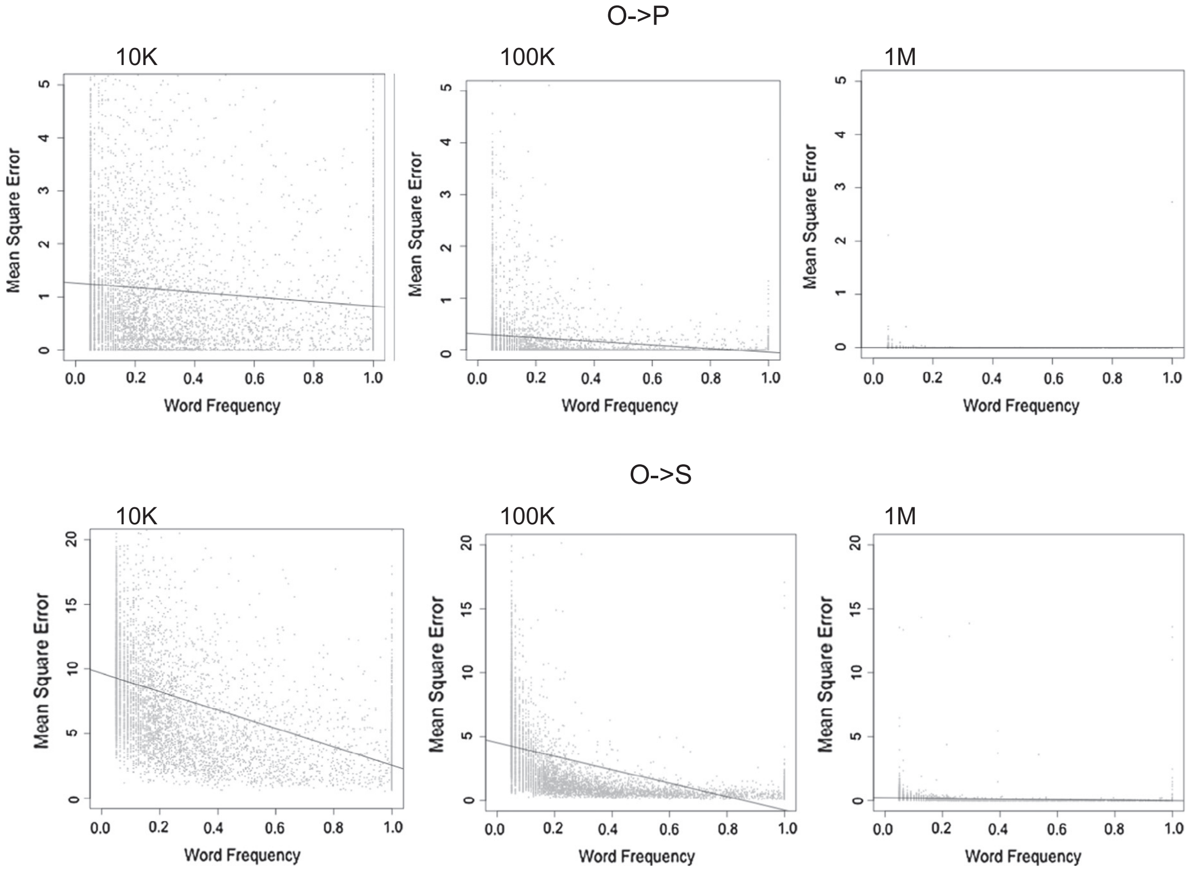

Fig. 5. Mean square error of the model's productions by word frequency for all 6229 words in the vocabulary, for orthography to phonology (O → P) and orthography to semantic (O → S) mappings at different stages of training. Solid lines show the linear regression fit.

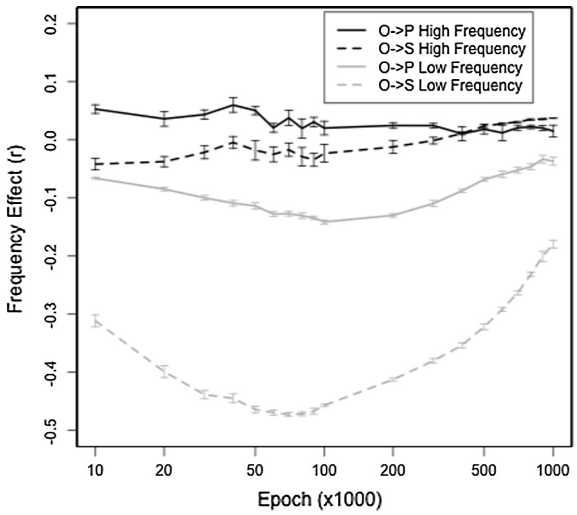

Fig. 6. Frequency effect for orthographic to phonological (O → P) and orthographic to semantic (O → S) representations across learning for higher and lower frequency words. Error bars show ±1 SEM of mean correlation between word frequency and error by simulation.

a response. When mean and variance associated with lexical access reduce in absolute terms (as for our model after extended training), the variation in responses associated with lexical access will be overwhelmed by noise associated with decision making processes (Gomez & Perea, 2014; Norris, 2009; Ratcliff, Gomez, & McKoon, 2004). Thus, absolute error reduction in lexical access will further reduce observations of the frequency effect, as a consequence of exposure, in behavioural studies of reading.

Simulation 2: frequency effects in the triangle model trained with varying vocabulary size

In Simulation 1 there was a confound between the size of the triangle model's vocabulary and the size of the frequency effect in terms of the model's performance: The model developed good representations for more words as training increased. Thus, it is not possible from Simulation 1 to discern whether the reduction in the frequency effect was directly caused by increased exposure to vocabulary, or whether the effect of exposure on frequency effects was mediated by increasing vocabulary knowledge. In order to determine whether the frequency effect was dependent on the model's vocabulary size, we repeated the simulations but varied the size of the vocabulary that the model learned. If the frequency effect was found to reduce as a consequence of vocabulary knowledge then a smaller trained vocabulary should result in a larger frequency effect than found in Simulation 1. If, however, the frequency effect is due to exposure then the vocabulary size of the training set should not affect performance, but rather the frequency effect should be related directly to the number of exposures to words.

Method

Architecture

The architecture was the same as in Simulation 1.

Training and testing

We compared performance of the model when trained on 1000, 2000, and 4000 words. The set of words used for training each vocabulary size was selected at random from the 6229 words used in Simulation 1. The smaller word sets were used for both the pretraining between phonological and semantic representations, and for the full triangle reading model. Pretraining on phonology to and from semantics was stopped after 2 million patterns had been presented. Training on the reading task was stopped after 1 million training trials.

The subsets of words were randomly selected for each of five separate runs of the model, to minimise differences in performance associated with particular random subset selections from the vocabulary.

Results

The triangle model's performance was assessed in the same way as for Simulation 1 by constructing mixed effects models and testing individual factors and interactions for their improvement to model fit.

Fig. 7 shows the accuracy of the model for mapping from orthography to phonology and orthography to semantics during learning for the 1000, 2000, and 4000 word sets. As anticipated, increasing the size of the vocabulary resulted in a reduction in accuracy during training: A generalized linear mixed effects model adding vocabulary size as a fixed factor improved model fit compared to a model with just random effects of simulation and word and fixed effect of log of training epoch, $\chi^2(1) = 31{,}081$, $p < .001$. This effect of vocabulary size was greater for mapping to semantics than mapping to phonology: Adding an interaction between mapping and vocabulary size increased model fit compared to the model containing just main effects, $\chi^2(1) = 4957.4$, $p < .001$. The effect of vocabulary size was particularly observed in the earlier stages of training: adding an interaction between log of epoch training and vocabulary size increased model fit further, $\chi^2(1) = 1090.4$, $p < .001$.

Fig. 8 shows the frequency effect for the model during training on different sized word sets, for both the semantic and the phonological output representations. To compare frequency effects for phonological and semantic representations, we found that, as for Simulation 1, adding the interaction between frequency and mapping improved fit, $\chi^2(1) = 48{,}281$, $p < .001$. Thus, the frequency effect was larger for the semantic representations than for the phonological representations. However, the interaction between frequency, mapping, and vocabulary size also improved

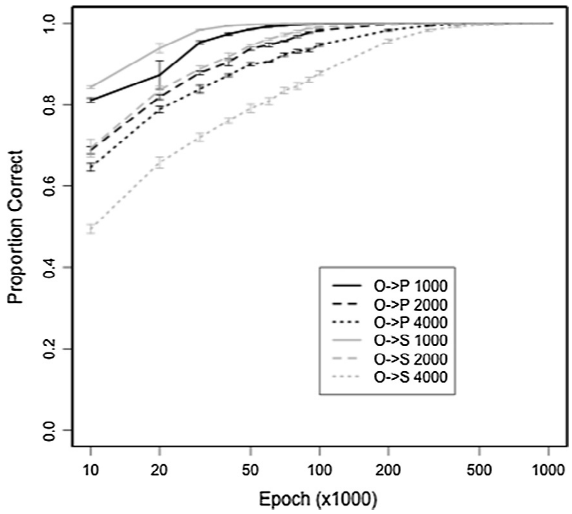

Fig. 7. Orthography to phonology and orthography to semantics mappings accuracy for the model trained with different vocabulary sizes. Overall, the larger vocabulary models performed less well. This was particularly true for orthography to semantics.

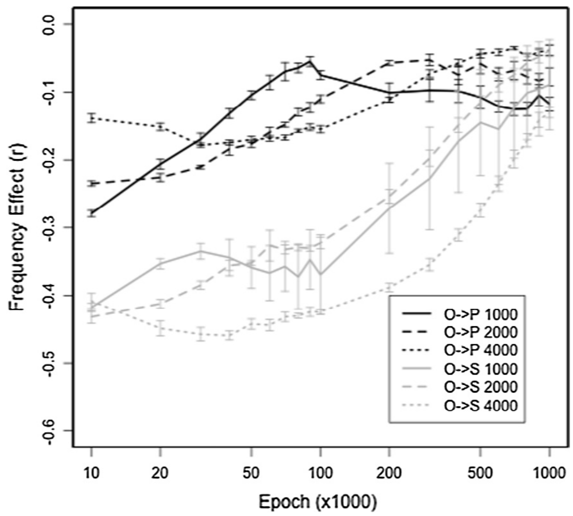

Fig. 8. Frequency effects for orthography to phonology and semantics mappings, for the triangle model trained with different vocabulary sizes. Error bars show ±1 SEM of mean correlation between word frequency and error by simulation.

fit, $\chi^2(1) = 8379.3$, $p < .001$. The difference in frequency effect between phonological and semantic representations was greater for the 4000 vocabulary than the 2000 vocabulary, $t = 66.6$, and the 2000 vocabulary resulted in a greater difference than the 1000 vocabulary, $t = 23.9$, both $p < .001$. The larger vocabulary size resulted in a greater difference in difficulty of learning arbitrary (semantic) versus quasi-systematic (phonological) mappings because learning a larger set of arbitrary patterns is more difficult than learning a smaller set (as shown in Fig. 7).

For each vocabulary size, the model demonstrated a reduction in the frequency effect during training. We tested the effect of vocabulary size on the frequency effect for the phonological and semantic representations separately, by first constructing a baseline linear mixed effects model with the closeness of the model's production to the target as the dependent variable, random effects of simulation and word, and fixed effects of log of training epoch, frequency and vocabulary size. The effect of vocabulary size on the frequency effect is determined by examining the interactions between the fixed effects.

For the phonological representations, adding the interaction between frequency and log of training epoch resulted in a significant improvement in fit, $\chi^2(1) = 11{,}711$, $p < .001$, thus confirming the effect of frequency changing with training that was also observed for the full set of 6229 words. Adding the interaction between frequency and vocabulary size significantly improved model fit, $\chi^2(1) = 266.13$, $p < .001$, with the magnitude of the frequency effect greater for 4000 words than 2000 words, $t = 6.38$, and the frequency effect for 2000 words greater than that for 1000 words, $t = 12.09$, both $p < .001$. Adding the three-way interaction between log of training epoch, frequency and vocabulary size to a model with all main effects and two-way interactions also resulted in a significant improvement in fit, $\chi^2(1) = 4.5263$, $p = .034$. The decline in the frequency effect with training was greater for the 2000 word vocabulary than the 4000 word vocabulary, $t = 7.11$, and the 4000 vocabulary decline was greater than the 1000 word vocabulary, $t = 8.64$, both $p < .001$. Thus, the change in the frequency effect was affected by vocabulary size, but was not monotonically related to vocabulary size: a larger vocabulary resulted in a *smaller* reduction in the frequency effect than a medium vocabulary. Overall, controlling for vocabulary size, the observation that frequency effects declined with training exposure was highly reliable.

We further tested whether the observation from Simulation 1 that the frequency effect changed direction as a consequence of training for the varying vocabulary sizes. We compared models with a linear and a quadratic interaction effect of frequency and log epoch, and found that the quadratic improved fit of the model over all three vocabulary sizes combined, $\chi^2(2) = 30{,}394$, $p < .001$, indicating that, overall, there was a quadratic effect of frequency against exposure similar to Simulation 1. However, the three-way interaction between vocabulary size, frequency, and quadratic function of log epoch improved fit further, $\chi^2(1) = 7312.1$, $p < .001$, indicating that the quadratic effect decreased with smaller vocabulary sizes. Investigating the vocabulary sizes individually, the quadratic effect improved model fit for all vocabulary sizes: for 1000 words, $\chi^2(2) = 7974.7$; for 2000 words, $\chi^2(2) = 12{,}300$; for 4000 words, $\chi^2(2) = 15{,}510$, all $p < .001$. Though Fig. 8 illustrates an initial increase for the 1000 word vocabulary for phonological representations, the quadratic fit indicates that the change in direction occurs at an early point in training. Thus, the change of direction in the frequency effect is greater for larger vocabulary sizes, but the effect is still discernible for smaller vocabulary sizes. We interpret this as being due to the difficulties in developing high-fidelity representations when the vocabulary size is greater, resulting in a larger initial increase in frequency effects with the larger vocabularies.

For the semantic representations, the same series of models were tested as for the phonological representations. The interaction between frequency and log of training epoch improved model fit significantly, $\chi^2(1) = 58{,}475$, $p < .001$. Frequency by vocabulary size also improved model fit, $\chi^2(1) = 23{,}879$, $p < .001$, with the frequency effect largest for 4000 words, then 2000 words, then 1000 words, $t = 73.2$, $t = 29.1$, both $p < .001$. Adding the three way interaction also significantly improved fit, $\chi^2(1) = 3851.7$, $p < .001$. In this case, there was a monotonic relation between vocabulary size and change in the frequency effect, such that the rate of change was highest for 4000 words than 2000 words, which was in turn higher than for 1000 words, $t = 36.60$, $t = 28.36$, respectively, both $p < .001$. However, importantly it remained the case that, when controlling for vocabulary size, frequency effects reduced as exposure increased.

The change in frequency effect with exposure was again found to be improved by a quadratic fit over the three vocabulary sizes, $\chi^2(2) = 20{,}977$, $p < .001$, however, as with the phonological representations, the interaction between vocabulary size and frequency and the quadratic of log epoch also significantly improved fit, $\chi^2(2) = 11{,}275$, $p < .001$. For each vocabulary size individually, the quadratic improved fit: 1000 words: $\chi^2(2) = 11{,}834$; 2000 words: $\chi^2(2) = 17{,}899$; 4000 words: $\chi^2(2) = 8657.6$, all $p < .001$. All vocabulary sizes demonstrated the change in direction of the frequency effect, though this was largest for the 2000 word condition.

All in all, there is little evidence that larger vocabulary sizes lead to smaller frequency effects. If anything, they induce stronger overall frequency effects. Furthermore, at least in the case of orthography to phonology mappings, a larger vocabulary is even protective against a change in frequency effects as a consequence of additional training. Thus, the behavioural effects relating to frequency effect changes are not simulated in the model by vocabulary size increasing, but are due instead to exposure. Furthermore, our interpretation of the frequency effect change as being driven by two processes – an initial increase in the frequency effect as representational fidelity improves, then decrease with exposure to items – is shown to be generalizable across these vocabulary sizes.

However, in Simulation 2 the selection of subsets of words was random which may not perfectly reflect the situation of actual acquisition, where smaller vocabularies are likely to comprise the most frequent words. In order to test whether vocabulary size might affect frequency effects if smaller vocabularies constitute the subset of higher-frequency words, we conducted Simulation 3.

Simulation 3: frequency effects in the triangle model trained with varying vocabulary size

This simulation was similar to that of Simulation 2, except that the subsets of 1000, 2000, and 4000 words comprised the most frequent words from the larger vocabulary, in order to simulate the greater likelihood of smaller vocabularies being composed of higher frequency words. We predicted similar effects to those of Simulation 2, namely that we would observe a reduction in the frequency effect for all vocabulary sizes with training, and that a larger vocabulary size would not relate to reduced frequency effects than a smaller vocabulary.

Method

Architecture

The architecture was the same as in Simulation 1.

Training and testing

We compared performance of the model when trained on the 1000, 2000, or 4000 most frequent words from the 6229 words used in Simulation 1. Training and testing was otherwise identical to that of Simulation 2.

Results

The triangle model's performance was assessed in the same way as for Simulation 2 by constructing mixed effects models and testing individual factors and interactions for their improvement to model fit.

Fig. 9 shows the accuracy of the model for mapping from orthography to phonology and orthography to semantics during learning for the 1000, 2000, and 4000 word sets. As for Simulation 2, increasing the size of the vocabulary resulted in a reduction in accuracy during training: A generalized linear mixed effects model adding vocabulary size as a fixed factor improved model fit compared to a model with just random effects of simulation and word and fixed effect of log of training epoch, $\chi^2(1) = 6514.8$, $p < .001$. Again, like Simulation 2, the effect of vocabulary size was significantly different for mapping to semantics than mapping to phonology: Adding an interaction between mapping and vocabulary size increased model fit compared to the model containing just main effects, $\chi^2(1) = 1149.8$, $p < .001$. Also similar to Simulation 2, the effect of vocabulary size was greater in the earlier stages of training: adding an interaction between log of epoch training and vocabulary size significantly increased model fit, $\chi^2(1) = 1568.6$, $p < .001$.

Fig. 10 shows the frequency effect for the model during training for semantic and the phonological output for the different vocabulary sizes in Simulation 3. There was a reduction in the frequency effect as vocabulary size reduced. As Simulation 2, the effect of vocabulary size on the frequency effect was determined by examining the interactions between the fixed effects, by testing the improvement of fit over a baseline model containing only random effects and main effects.

As for Simulations 1 and 2, frequency effects were found to be larger for semantic than phonological representations, $\chi^2(1) = 41{,}545$, $p < .001$. As for Simulation 2, the interaction between frequency, mapping, and vocabulary size improved fit, $\chi^2(1) = 5197.5$, $p < .001$. The difference in frequency effect was greater for 4000 than 2000 words, $t = 60.3$, and greater for 2000 than 1000, $t = 6.0$, both $p < .001$, consistent with an enhanced difference for a model required to learn a larger versus a smaller set of

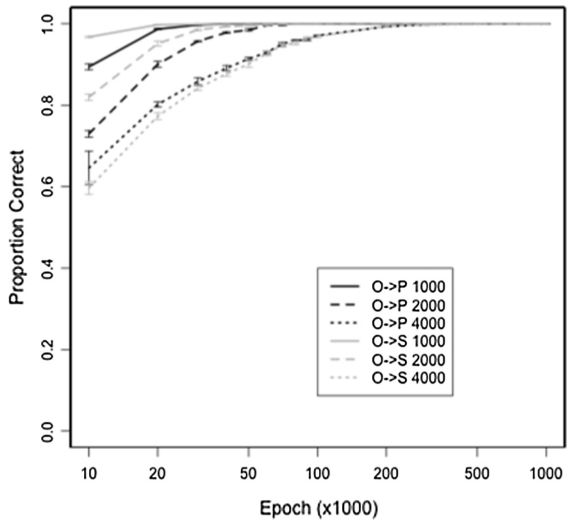

Fig. 9. Orthography to phonology and orthography to semantics mappings accuracy for the model trained with different vocabulary sizes, selected as the most frequent words.

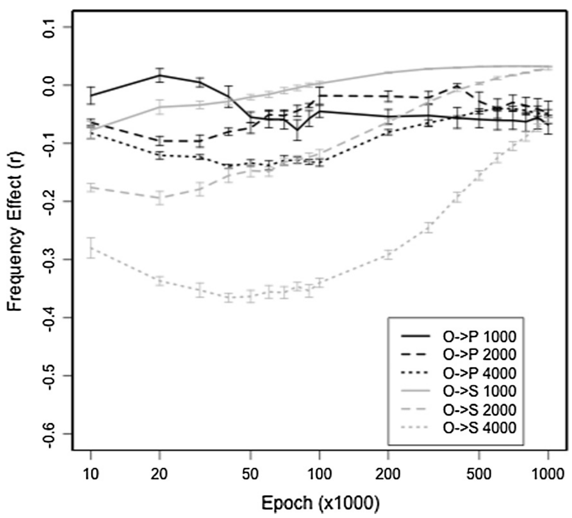

Fig. 10. Frequency effects for orthography to phonology and semantics mappings, for the triangle model trained with different vocabulary sizes for the most frequent 1000, 2000, or 4000 words in the corpus. Error bars show ±1 SEM of mean correlation between word frequency and error by simulation.

harder arbitrary (semantic) versus easier quasi-systematic (phonological) mappings.

For the phonological representations, adding the interaction between frequency and log of training epoch resulted in a significant improvement in fit, $\chi^2(1) = 12{,}613$, $p < .001$, the effect of frequency changed with training in the same way as for Simulations 1 and 2. Adding the interaction between frequency and vocabulary size significantly improved model fit, $\chi^2(1) = 33.037$, $p = .001$, with the magnitude of the frequency effect greater for

4000 words than 2000 words, $t = 2.31$, $p = .021$, which was greater than 1000 words, $t = 2.52$, $p = .012$. Adding the three-way interaction between log of training epoch, frequency and vocabulary size to a model with all main effects and two-way interactions also resulted in a significant improvement in fit, $\chi^2(1) = 176.93$, $p < .001$. The effects were similar to those for Simulation 2: the larger vocabulary related to a larger frequency effect. When vocabulary size was controlled, the frequency effect was found to decrease as a consequence of extended training.

As for Simulation 2, the change in frequency effect with exposure was found to be improved by a quadratic fit over the three vocabulary sizes, $\chi^2(2) = 46{,}623$, $p < .001$. Also as for Simulation 2, the interaction between vocabulary size, frequency and quadratic of log epoch also improved fit, $\chi^2(2) = 32{,}477$, $p < .001$. For each vocabulary size individually, the quadratic again improved fit: 1000 words: $\chi^2(2) = 5098.5$; 2000 words: $\chi^2(2) = 20{,}332$; 4000 words: $\chi^2(2) = 26{,}010$, all $p < .001$. Again, all vocabulary sizes demonstrate the change in direction of the frequency effect.

For the semantic representations, the interaction between frequency and log of training epoch improved model fit significantly, $\chi^2(1) = 63{,}787$, $p < .001$. Frequency by vocabulary size also significantly improved model fit, $\chi^2(1) = 4.927$, $p = .026$. Adding the three way interaction did significantly improve fit, $\chi^2(1) = 4268.9$, $p < .001$. As with the phonological effects, the larger vocabularies resulted in a larger frequency effect, and demonstrated that, when controlling for vocabulary size, the frequency effect reduced with exposure.

For the quadratic fit of log epoch, the interaction with frequency improved fit over all three vocabulary sizes for the semantic representations, $\chi^2(2) = 56{,}523$, $p < .001$. There was a significant improvement in fit with the interaction between vocabulary size, frequency and the quadratic of log epoch, $\chi^2(2) = 37{,}148$, $p < .001$. As with Simulation 2, the quadratic improved fit for each vocabulary size: 1000 words: $\chi^2(2) = 11{,}834$; 2000 words: $\chi^2(2) = 17{,}899$; 4000 words: $\chi^2(2) = 8657.6$, all $p < .001$. The results show that, as for Simulation 2, there is a change in direction of the frequency effect with training, with the size of the effect changing, but the qualitative nature of this change unaffected by vocabulary size.

Thus, the results of Simulation 2 and 3 indicate that a larger vocabulary was protective against reduced frequency effects, rather than the cause of frequency effect changes with training as could be expected given the stronger competition possible from a larger vocabulary. Therefore, the smaller frequency effect for people with large vocabularies found in lexical decision tasks cannot be explained by vocabulary size itself. At the same time, Simulations 2 and 3 confirmed the finding of Simulation 1 that extra exposure undoes the larger frequency effect related to the knowledge of more words. Towards the end of the training, the frequency effect was similar for all vocabulary sizes tested. After 1 million training trials the frequency effect on the O → S mappings was smaller for the model trained on 4000 words than for the model with 1000 words trained after 20K trials, even when the latter 1000 words were the most frequent ones (Fig. 10).

Simulation 4: frequency effects in first and second languages

Simulations 1, 2, and 3 established that, in the triangle model, the frequency effect in learning to read a single language can relate to exposure. In bilinguals, mapping between orthographic, phonological, and semantic representations in two languages, frequency effects have been shown to be stronger compared to monolinguals (Gollan, Montoya, Cera, & Sandoval, 2008; Ransdell & Fischler, 1987). An explanation for this has been in terms of frequency of usage (Gollan et al., 2008): As bilinguals have less exposure to each language, they have "weaker-links" between orthographic, phonological, and semantic representations and this will be particularly harmful for accessing low frequency words.

An alternative account of reduced frequency effects is increased interference between languages: there is greater competition amongst a vocabulary that is almost twice as large in bilinguals than monolinguals, reducing the psycholinguistic effects influencing lexical access in a single language (Costa, 2005; Peterson & Savoy, 1998). Such influences across languages are well-attested, with L2 acquisition resulting in slower lexical access to L1 (Kroll, Michael, Tokowicz, & Dufour, 2002; Linck, Kroll, & Sunderman, 2009) and a larger frequency effect, even in the dominant language (Gollan et al., 2008).

In terms of comparison of frequency effects within bilingual speakers, the frequency effect is typically larger in L2 than in L1 (Cop, Keuleers, Drieghe, & Duyck, 2015; de Groot, Borgwaldt, Bos, & van den Eijnden, 2002; Duyck, Vanderelst, Desmet, & Hartsuiker, 2008; Van Wijnendaele & Brysbaert, 2002; Whitford & Titone, 2012). In a mega-study, Lemhöfer et al. (2008) tested English word identification in English monolingual and bilingual Dutch, French, and German speakers, and found a larger L2 frequency effect than L1 in English, which was due principally to greater slowing of low-frequency words in the L2 speakers. Diependaele et al. (2013) argued that this difference disappears when vocabulary size in each language is taken into account, and in a more recent mega-study Brysbaert et al. (in press) confirmed that most of the difference in frequency effects between L1 and L2 was due to vocabulary size, taken as a proxy for exposure to each language.

In the present simulation, we investigated whether the triangle model can simulate these effects by examining relative exposure to two languages. We tested two hypotheses: (1) that exposure is the main determinant of the difference in frequency effect between L1 and L2, and (2) that knowledge of another language increases the frequency effect in L1. We tested whether these hypotheses were consistent with the triangle model's performance when trained on a second language. We chose Dutch as the second language, as this language has a high degree of orthographic overlap with English and was one of the languages tested by Diependaele et al. (2013). We implemented sequential acquisition (L2 introduced after some time learning L1), as this is the typical state-of-affairs for participants in research on bilingualism (Li & Zhao, 2013).

Method

Architecture

The architecture of the model was the same as in Simulation 1.

Training and testing

The model was initially trained in the same way as Simulation 1, with pretraining on the 6229 English words, between semantics and phonological representations, and then training for 500,000 presentations of English orthographic words mapping onto phonology and semantics.

At this point the model was then trained on learning to read an expanded set of words, including all 6229 words in English together with 1536 Dutch words. These Dutch words were all the words with the same meaning as English which were translated using Google translate as individual monosyllabic words, and which were not identical to the English orthographic forms. The orthographic forms for these Dutch words were taken directly from the translation in Google translate. Phonological forms were derived using the Dutch CELEX database, with Dutch vowels mapped onto the nearest English noun, e.g., Dutch /ɑ/ (as in bad /bɑt/) became English /æ/ (as in hat /hæt/), and Dutch consonants mapping to the nearest English consonant, e.g., Dutch /ʋ/ (as in wat /ʋɑt/) became English /w/ (as in wit /wIt/), with the exception of /x/ which was added to the phonological inventory of the model. The semantic forms were the same as the yoked English words. Frequency was also taken to be the same as the yoked English form. Thus, the model learned to map from new orthographic forms to new phonological forms, but onto previously experienced semantic representations.

The model was trained to read for a further 500,000 presentations of orthographic words mapping to phonology and semantics. Words were selected randomly according to frequency from the combined set of English and Dutch words. Based on these frequencies, approximately 25% of patterns were Dutch and the remainder English words.

Additional simulations tested the effect of the varying proportion of English and Dutch words by increasing the relative frequency of the Dutch words. We investigated simulations where 50% of training patterns were English and 50% were Dutch, and simulations where 25% of patterns were English and 75% were Dutch.

The model was tested in the same way as in Simulation 1, except that the test set included all English and all Dutch words experienced during training.

Results

Fig. 11 shows the model's learning from the original monolingual simulation (Simulation 1) against the model's performance for learning both English and Dutch from 500,000 patterns onwards, after initial training only on English. For comparison to the original simulation, performance for the 75% English and 25% Dutch simulation is shown.

For English words in the bilingual model, there was a slight decrement in performance that was caused by introduction of the Dutch words. As Dutch requires some remapping of the orthography to phonology statistical relations, and interleaving of new orthography to semantics representations, this caused interference in the model's performance. For instance, for orthographic input "bad", the model has to map onto the phonology of both /bæt/ in Dutch and /bæd/ in English, and onto meanings of "bath" in Dutch and "bad" in English. Furthermore, mapping onto the semantics of "bath" would also result in interference along the semantics to phonology pathway, as the model learns that this can map onto the phonological form in English or in Dutch. However, performance remained highly accurate and recovers to levels close to perfect performance at the end of training.

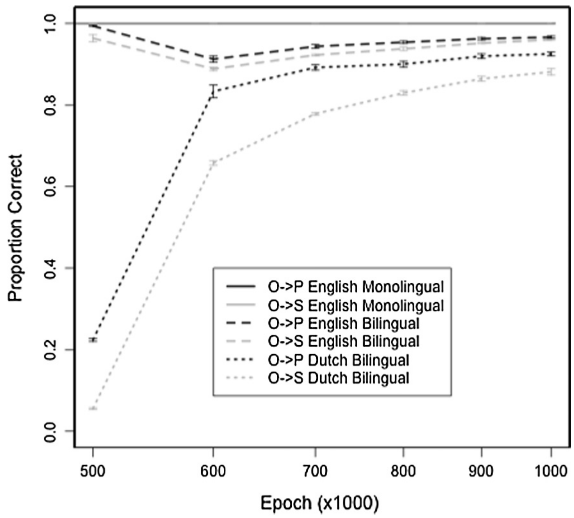

Fig. 11. Performance of the English monolingual model (Simulation 1), and the bilingual model on English and on Dutch for mapping from orthography to phonology and to semantics. Note that the English monolingual orthography to phonology and orthography to semantics are superimposed.

Interestingly, for the Dutch words, before training on any Dutch mappings the triangle model is able to accurately read some Dutch words. This is not so surprising for the orthography to phonology mappings because there is considerable overlap in the letter-to-sound correspondences in English and Dutch, and so the model is able to generalise accurately to approximately 24% of the Dutch words. However, the model is also able to correctly generalise to approximately 6% of orthography to semantics mappings in Dutch. This is somewhat surprising because perfect cognate forms were not included in the Dutch word set, but it indicates that very similar orthographic forms had, in these cases, similar meanings in Dutch and in English (e.g., Dutch bal /bɑl/, English ball /bɔ:l/, Dutch licht /lIxt/, English light /laIt/). Over time, the triangle model's learning of the Dutch mappings improves, most rapidly for phonology, and more gradually for Dutch meanings. Overall, then, the model was effective at learning to read bilingually: by the end of training the triangle model had high accuracy in mapping words in both English and Dutch.

The effect of the relative frequency of English and Dutch words during training on learning the orthography to phonology mappings is shown in Fig. 12 for English words, and Fig. 13 for Dutch words. Unsurprisingly, increased exposure in Dutch resulted in increased speed of learning for Dutch. The fit of a generalized linear mixed effects model on accuracy for phonological representations in Dutch, with simulation and word as random effects, and training epoch as fixed effect was significantly improved by adding proportion of Dutch training, $\chi^2(1) = 1144$, $p < .001$. Accuracy was higher for 75% Dutch than 50% Dutch, $t = 12.89$, which was in turn higher than 25% Dutch, $t = 20.98$, both $p < .001$. For English, decreasing exposure to English resulted in a smaller effect, but still reliable impact on learning, $\chi^2(1) = 848.43$, $p < .001$, with lower accuracy for 75% Dutch than 50% Dutch in English reading, $t = 15.59$, and 25% Dutch exposure resulted in still higher English reading accuracy, $t = 13.58$, both $p < .001$.

For orthography to semantic mappings, proportion of Dutch exposure again had an influence on Dutch reading accuracy, $\chi^2(1) = 2846.2$, $p < .001$, with 75% Dutch exposure resulting in higher accuracy than 50% Dutch exposure, which was in turn more accurate than 25% Dutch exposure, $t = 14.18$, $t = 37.05$, respectively, both $p < .001$. For English reading accuracy, exposure was again a significant factor, $\chi^2(1) = 1877.6$, $p < .001$, with 75% Dutch exposure resulting in lower accuracy than 50%, $t = 17.92$, which was in turn lower than 25% Dutch exposure, $t = 25.68$, both $p < .001$.

The effect of varying exposure to a second language on frequency effects in first and second language is shown in Fig. 14.

For English, increased exposure to the second language (Dutch) resulted in increased frequency effects in English. A linear mixed effects model on closeness of the model's phonological production to the target phonology as dependent variable, simulation and word as random effects, log epoch of training, frequency, and proportion of Dutch training as fixed effects was improved in fit by adding an interaction between frequency and proportion of Dutch training, $\chi^2(1) = 72.014$, $p < .001$, with 75% Dutch exposure resulting in a larger magnitude of the frequency effect than 50% Dutch exposure, $t = 8.50$, which was in turn larger than the effect from 25% Dutch exposure, $t = 5.51$, both $p < .001$. The rate of change of the frequency effect with length of exposure was also related to the proportion of Dutch training. Adding the three-way interaction between frequency, log of epoch of training, and proportion of Dutch exposure to a model containing random effects and main and two-way effects improved model fit significantly, $\chi^2(1) = 7.350$, $p = .007$. The rate of change was greater for 75% exposure to Dutch than 50% exposure to Dutch, $t = 2.71$, $p < .001$, but 50% and 25% exposure to Dutch did not differ, $t = .23$, in their effects on the frequency effect in English (see Fig. 14A).

For semantic representations in English, the effect of Dutch exposure was greater still. The interaction between frequency and proportion of Dutch exposure significantly improved model fit, $\chi^2(1) = 133.8$, $p < .001$, with 75% Dutch exposure resulting in a greater magnitude of the frequency effect than 50% Dutch, $t = 11.62$, which was greater in turn than 25% Dutch, $t = 16.02$, both $p < .001$. The three-way interaction also significantly improved model fit, $\chi^2(1) = 11.217$, $p < .001$, with the rate of change greater for the 75% Dutch exposure than 50%, $t = 3.36$, which was greater rate of change than for the 25% exposure, $t = 4.95$, both $p < .001$ (Fig. 14C).

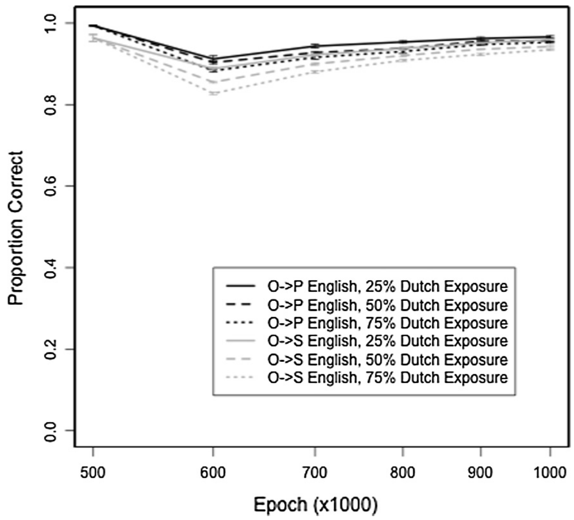

Fig. 12. Effect of varying exposure to Dutch on English reading accuracy.

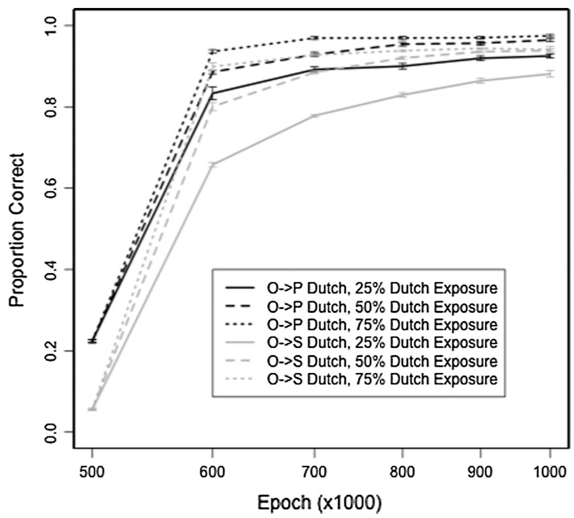

Fig. 13. Effect of varying exposure to Dutch on Dutch reading accuracy.

Increase in exposure to the second language (Dutch) resulted in a decrease in frequency effects in second language (Dutch) for phonological representations: a linear mixed effects model on the closeness of the model's phonological productions to the target with simulation and word as random effects and word frequency and proportion of Dutch as main effects was improved in its fit by adding the interaction between frequency and proportion of Dutch exposure, $\chi^2(1) = 98.59$, $p < .001$. 25% Dutch exposure resulted in a higher frequency effect than 50%, $t = 23.16$, which was in turn higher than 75%.

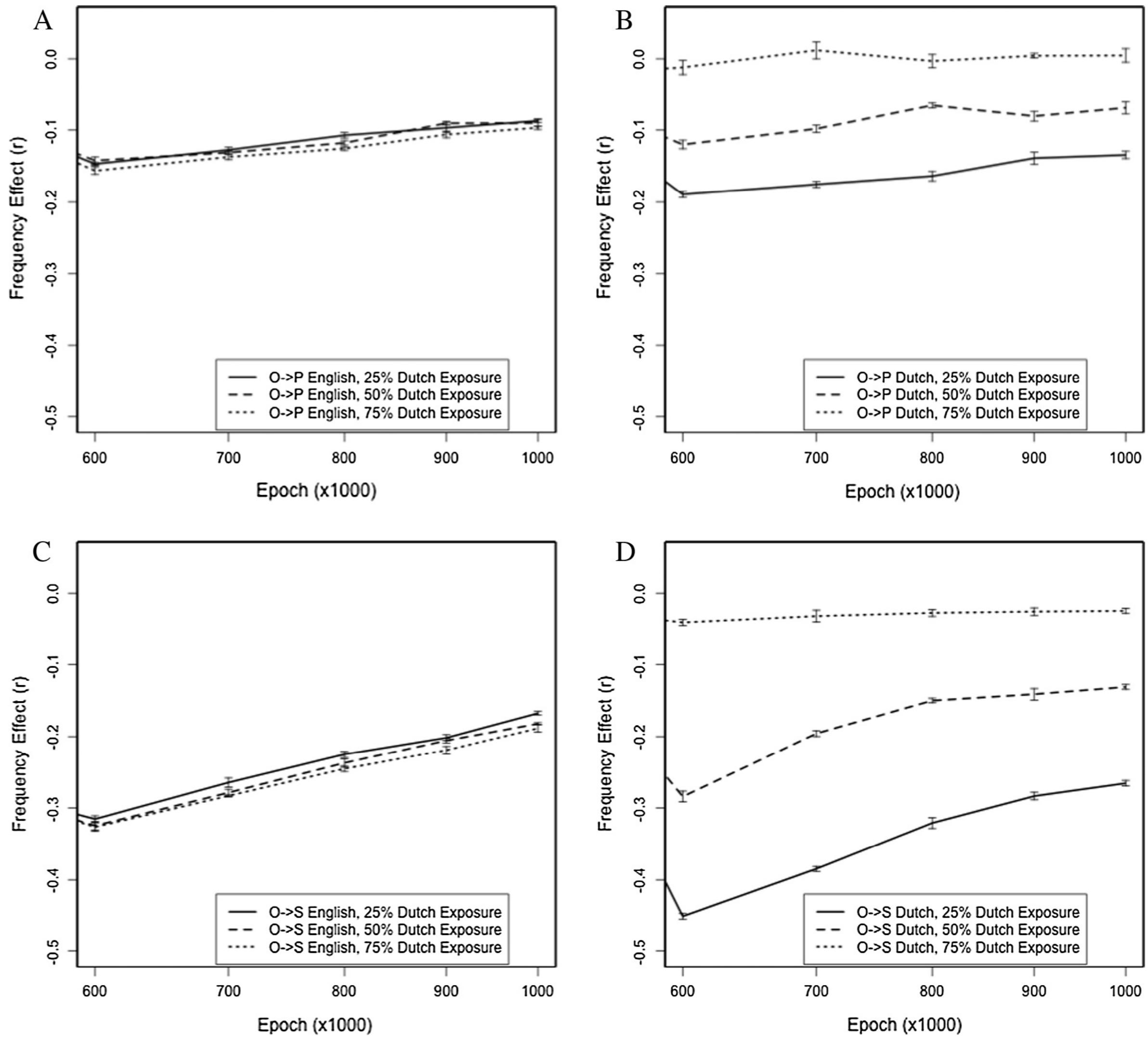

Fig. 14. Frequency effect affected by exposure to second language. (A) Effect of Dutch exposure on English orthography to phonology; (B) effect on Dutch orthography to phonology; (C) effect on English orthography to semantics; (D) effect on Dutch orthography to semantics. Notice that as the curves go higher in this figure, they approach a frequency effect of 0; lower values mean a stronger frequency effect.

Dutch exposure, $t = 14.33$, both $p < .001$. Adding the interaction between frequency, Dutch exposure, and log epoch training exposure improved fit compared to a model containing main effects and two-way effects, $\chi^2(2) = 68.286$, $p < .001$, indicating a greater change of frequency effect with training time for the 25% Dutch exposure than 50% Dutch exposure, $t = 3.311$, and smaller change still for the 75% Dutch exposure, $t = 4.902$, both $p < .001$ (see Fig. 14B).

For semantics, increase in exposure to Dutch also resulted in a reduced effect of frequency for Dutch (Fig. 14D). Adding the interaction between word frequency and proportion of Dutch improved model fit, $\chi^2(1) = 899.47$, $p < .001$, with the magnitude of the effect significantly larger for 25% Dutch than 50% Dutch exposure, $t = 41.09$, and smaller still for 25% Dutch, $t = 31.08$, $p < .001$. Adding the three-way interaction to the model also improved fit, $\chi^2(1) = 118.17$, $p < .001$, indicating that the effect declined at a greater rate with further training for 25% compared to 50%, $t = 8.60$, and 50% compared to 75% Dutch exposure, $t = 11.28$, both $p < .001$.

All in all, the results of the simulations agree rather well with the behavioural findings: (1) The English frequency effects become stronger with more use of Dutch, but (2) decrease as the training continues. We also see (3) a stronger frequency effect in L2 when it is used less frequently (i.e., for the Dutch 25% exposure). However, from Fig. 14, there is a suggestion that the frequency effect in the 75% Dutch condition was very small (panels B and D). This was partially a consequence of measuring the frequency effect only after 100,000 training presentations to the model. Comparing to the different vocabulary conditions of Simulation 2 (Fig. 8), after 100,000 epochs the frequency effect for phonological and semantic representations in the

1000 word simulation had already substantially declined. Investigating the Dutch model at earlier training stages, we found that frequency effects were initially higher than those observed after 100,000 bilingual training trials for the semantic representations: after 540,000 trials, the frequency effect peaked at −.093 (*SD* = .020). Yet, the frequency effect for phonology remained small, but significantly different than chance, at these earlier training stages (e.g., after 560,000 trials, the frequency effect peaked at mean = −.025, *SD* = .024). It could be that the small frequency effect in Dutch was due to optimising the merging of the statistics of the mappings for Dutch and English orthography to phonology mappings when sufficient exposure to Dutch was available, thereby resulting in Dutch words being processed with similar levels of ease regardless of their individual frequencies. As the overlap between orthography and semantics is only very low between these languages, we do not observe a reduced frequency effect for the semantic representations.

However, note that the simultaneous exposure to the two languages exacerbates the frequency effect: the 25% Dutch exposure model has had the same exposure to Dutch at 700,000 epochs of training as the 50% Dutch exposure model has had at 600,000 epochs, and yet the frequency effects appear to still be enhanced in this second language. To test this possible enhancement from learning in another language, independent of exposure to the language in which frequency effects are to be tested, we anal-

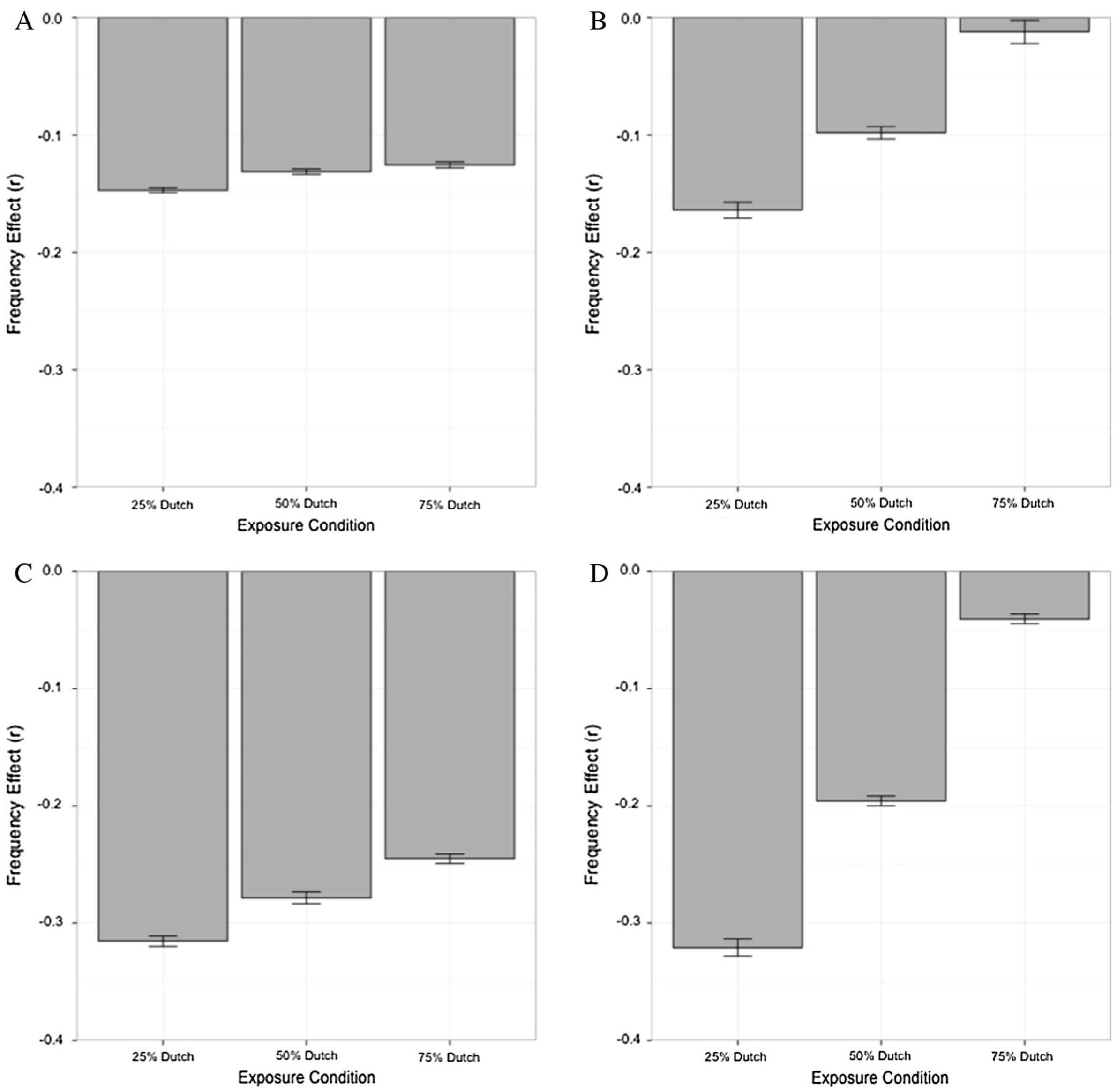

Fig. 15. Frequency effect according to exposure to second language, controlling for exposure in the first language. (A) English orthography to phonology; (B) Dutch orthography to phonology; (C) English orthography to semantics; (D) Dutch orthography to semantics.

ysed a subset of the model data equating the exposure to each language, and comparing the frequency effect across exposure conditions. Thus, for English, we compared the frequency effect of the model for the 25% Dutch exposure training at 600,000 epochs, the 50% Dutch exposure training at 700,000 epochs, and the 75% Dutch exposure training at 800,000 epochs. For Dutch, we measured the frequency effect for the 25% Dutch exposure training condition at 800,000 epochs, the 50% Dutch exposure training at 700,000 epochs, and the 75% Dutch exposure training at 600,000 epochs. The results for frequency effects in phonology and in semantics are summarised in Fig. 15.

Baseline linear mixed effects models on the frequency effect were first constructed, with simulation and word as random effects and frequency and exposure condition as factors. Then, the improvement in fit when the interaction between frequency and exposure condition was determined.

For English, the intensity of Dutch exposure had a significant effect for orthography to phonology mappings, $\chi^2(1) = 58.435$, $p < .001$, and for orthography to semantics, $\chi^2(1) = 811.99$, $p < .001$. Similarly, for Dutch, intensity of exposure was significant for orthography to phonology, $\chi^2(1) = 12.147$, $p < .001$, and for orthography to semantics, $\chi^2(1) = 83.464$, $p < .001$. The effects of intensity affected both languages in a similar way: there was greater reduction of the frequency effect if exposure to Dutch was more intense, which applied both to English words and Dutch words in the bilingual model.

A further analysis of the 25%, 50%, and 75% Dutch exposure simulations, controlling for accuracy of Dutch reading instead of exposure to Dutch, resulted in a similar pattern of effects. At 600,000 epochs, the 75% Dutch exposure simulations reached 93.6% (*SD* = 24.4%) for phonology and 89.9% (*SD* = 30.1%) for semantics. At 700,000 epochs, the 50% Dutch exposure simulations reached similar accuracy (phonology mean = 92.9%, *SD* = 25.7%; semantics mean = 88.5%, *SD* = 31.9%). At 1,000,000 epochs, the 25% Dutch exposure was similarly accurate (phonology mean = 92.5%, *SD* = 26.3%; semantics mean = 88.1%, *SD* = 32.4%), so these simulations at these training epochs were compared. Intensity of Dutch exposure influenced frequency effects in English for both orthography to phonology, $\chi^2(1) = 198.44$, $p < .001$, and orthography to semantics, $\chi^2(1) = 3241$, $p < .001$. Similarly, intensity of Dutch exposure influenced frequency effects in Dutch in phonology, $\chi^2(1) = 145.51$, $p < .001$, and semantics, $\chi^2(1) = 478.56$, $p < .001$. As with the simulations controlling for exposure to Dutch, when controlling for accuracy of performance in Dutch, increased intensity of Dutch exposure resulted in a smaller frequency effect in both languages.

Thus, frequency effects were not entirely independent in first and second language, and therefore cannot be completely accounted for by exposure within a language in the model's performance, as Diependaele et al. (2013) have proposed. Instead the results seem consistent with the weaker-links hypothesis of Gollan et al. (2008), who proposed that learning a second language can reduce the strength of mapping between orthography and phonology and semantics in a first language. This weaker links property of the model is an emergent result of training the model on multiple languages, and such weakening of links does not have to be explicitly included in the model.

General discussion

Individual differences in performance for language tasks are a topic of growing interest (Andrews, 2015; Yap et al., 2012). Such variation can provide insight into the processing parameters that underlie behaviour. In word naming and lexical decision tasks, a key observation is that psycholinguistic effects may vary across participants. Individual differences in the variance in response times and accuracy explained by psycholinguistic variables can be partially accounted for by age (Morrison, Hirsh, Chappell, & Ellis, 2002), by language proficiency (Chateau & Jared, 2000; Diependaele et al., 2013; Lewellen et al., 1993; Preston, 1935; Sears et al., 2008; Yap et al., 2008, 2012), or as a consequence of language exposure (Brysbaert et al., in press; Kuperman & Van Dyke, 2013). Of particular interest to us was to examine the potential causes of the frequency effect, because it accounts for a large proportion of variance in lexical processing accuracy and response times in behavioural studies. Our simulations were able to replicate observed differences in frequency effects for lexical processing tasks that principally involve mapping from orthography to phonology and those that map from orthography to semantics (Ghyselinck et al., 2004).

We considered four possible explanations for the observation that participants with larger vocabularies have lower frequency effects. First, the relation between size of the frequency effect and vocabulary size may be a mere side-effect of quicker response times in those with greater language proficiency. In this case, the frequency effect may be reduced in those with higher language proficiency because of a floor effect in response times. Our Simulation 1 demonstrated that greater proficiency could be related to frequency effects, but went further than previous behavioural studies by demonstrating a potential cause of this relation: due to amount of exposure to language by the reading system. Furthermore, the origin of the reduced frequency effect was primarily due to reduction of error variance for lower frequency words in the triangle model. This change in the model's mappings between representations is a consequence of error-driven learning in the model, such that those patterns that contribute most error contribute most change to weights on connections within the model. As error from low-frequency words is greater than that for high-frequency words, the low-frequency words are contributing most to reconfiguring the model's structure by reducing the model's error for those low-frequency patterns. Thus, the reduced frequency effect was not entirely due to a general improvement in response fidelity across all stimuli, in contrast to this first explanation. However, the overall reduction in error for the model's representations of phonological and semantic forms of words is consistent with a contribution of frequency effect reduction relating to response variation associated with psycholinguistic processes of lexical access associated with generation of the decision making response (e.g., Norris, 2009). Furthermore, we established in Simulations

2 and 3 that vocabulary size was not the key variable underlying changes in frequency effects, but rather amount of exposure was the critical driver behind efficient processing of mappings between representations.

The second explanation for variation in frequency effects was that language proficiency is related to intelligence, and intelligence is underwritten by greater speed of processing, which could again compress frequency effects for those with higher intelligence. Here, the data and computational modelling of bilingual participants is crucial. Diependaele et al. (2013) showed that frequency effects were not person-dependent but rather dependent on the individual's proficiency in the language being tested. We showed that the triangle model can be extended to learn to read words in second language, and that varying the exposure of the model to first and second language could predict the pattern of frequency effects for L1 and L2 speakers. Simulation 4 demonstrated that increased exposure to L2, with a concomitant increase in proficiency in L2, resulted in increased frequency effects in L1 and reduced frequency effects in L2. However, the model's performance was not wholly accounted for by amount of exposure within a language, as there was evidence that intensity of exposure also affected the size of frequency effects. In both first and second languages, greater intensity of second language exposure reduced the size of frequency effects when total exposure within each language was controlled. Indeed, increased intensity of exposure to a second language could be hypothesised to result in increasing the noise in mappings for the first learned language, thereby increasing the frequency effect in that first language, due to reduction in compression. However, the opposite was the case: the increase of the L1 frequency effect was largest in the 25% Dutch exposure situation. The finding that our model predicts an increase in the L1 word frequency effect when another language is learned, is a finding consistent with studies of interference effects across languages (e.g., Costa, 2005; Gollan et al., 2008; Linck et al., 2009), where acquisition of an L2 can increase response times in L1. Such effects may be observed at both the lexical access stage of language processing (as demonstrated in our model) as well as affecting decision making processes, as reflected in the subtle effects of L2 revealed in the diffusion model simulations of behavioural results in Brysbaert et al. (in press).

The third explanation for reduced frequency effects is that greater exposure to a language results in proportionally more exposure to lower frequency words (Kuperman & Van Dyke, 2013). However, the frequency compression used for sampling of input to the model meant that even all the low frequency words were highly likely to occur even in small samples. For instance, by 200,000 random samples, the point at which frequency effects tend to decrease in magnitude, 99.9% of words will have been sampled. Furthermore, sampled word frequencies at 100,000, 200,000, and 300,000 epochs were correlated at 1.00. Thus, sampling biases are not sufficient to explain the triangle model's performance.

The fourth explanation we considered was that exposure is the key factor underlying the relation between language proficiency and size of frequency effects between individuals. Training a model that learns to map between orthographic and phonological and semantic representations with increasing efficiency demonstrated the same effects as those observed in participants. Furthermore, size of vocabulary was not sufficient to explain the model's performance. The triangle model therefore tests the adequacy of a theory based on language exposure resulting in greater efficiency of accessing representations of words. This theoretical principle was shown to account also for individual difference effects observed in reading in L1 and L2, and these data are critical for distinguishing exposure effects from other individual variation in cognitive processing that could affect performance. For instance, efficiency of mappings between representations can be the result of the amount of resources serving mappings in a computational model, or by the learning function – faster learning relates to a higher learning rate parameter in the model, or by increasing the speed with which information can pass within the network (e.g., Faust et al., 1999; Plaut & Booth, 2000). All these parameters are potentially adjustable in the model, but none would explain the apparent interaction between size of frequency effects in L1 and L2. Adjustments to resources, rate of learning, or speed of processing would result in similar effects in both first and second language, whereas, the size of the frequency effects are shown to be inversely related to proficiency for each language. We thus contend that additional factors contributing to individual differences in the frequency effect are not necessary to explain the data, and that an explanation based on exposure is the most parsimonious explanation for the observed effects.

Similarities between first and second languages may influence the extent to which multiple languages influence processing in the other language. Kaushanskaya, Yoo, and Marian (2011) found that for English-Spanish bilinguals, proficiency in Spanish reading was associated with proficiency in English reading. However, for English-Mandarin bilinguals, self-reported Mandarin reading proficiency was associated with lower English reading skills. The simulations of bilingual reading we have performed have involved two closely-related languages, with overlapping orthographic and phonological mappings (consider the bad/bath example, above). In a behavioural study on naming responses in English, Lemhöfer et al. (2008) found only small differences in responses on English words between English monolingual, Dutch-English, French-English, or German-English bilinguals, apart from the enhanced frequency effect for L2 speakers. So, such closely-related languages may not result in a strong interference effect. Yet, simulating a wider range of languages, with varying degrees of similarity among orthographic, phonological, and semantic representations would enable us to determine the computational consequences of overlaying overlapping versus distinct mappings in the reading system.

Critically, the model predicted that changes in frequency effects were not linear as a consequence of exposure. Rather, frequency effects increased during early stages of language processing, as the model develops an accurate representation of words, and discrimination between phonological and semantic forms, akin to development of lexical quality (Perfetti, 2007). However, after these representations have become well-formed (from about 100,000 to 200,000 epochs of training) the frequency effect then begins to reduce, as a consequence of increasing efficiency of the mappings. Thus, the triangle model generates the prediction that individual differences in lexical processing are likely to reflect both this fidelity of representation and efficiency of mapping, and can potentially explain why frequency effects are less prominent in children than young adults (Ellis, 2002; Garlock, Walley, & Metsala, 2001), because frequency effects are reduced by poorer quality of representation. However, our simulations predict that with extensive exposure, frequency effects can in principle fall below those of learners in early stages of acquiring the language (see, e.g., Fig. 4), especially for lower-frequency words (Fig. 6). Comparisons between children and older adults would be one way to assess this prediction.

Adelman et al. (2014) examined a range of psycholinguistic factors, including length, consistency and frequency, in terms of parameter variation in DRC and CDP+ models. Their interest was the extent to which these models were sufficient to explain observed inter-individual covariation in psycholinguistic variables derived from behavioural mega-studies. Our aim for the current simulations was different: to distinguish the relative contributions of language proficiency and the size of the frequency effect in a computational model of reading that can learn mappings as a consequence of exposure to the vocabulary of a language. However, there are possibilities for investigating the extent to which variation in training the triangle model can reflect behavioural observations for other psycholinguistic variables. For instance, Adelman et al. (2014) co-located length and consistency effects in the sublexical route of the DRC and CDP+ models, and located frequency effects in the lexical route of these models. This constrains the extent to which these variables are likely to covary – length and consistency effects should have similar coefficients for individuals, but may have different coefficients to that of frequency. In the triangle model, we anticipate that variables such as length and consistency should be related to exposure in a similar way to frequency. This is because the effect of exposure on the model is to increase efficiency of the mappings, and compress the size of the difference between mappings that are initially difficult and those that are easier. Longer words tend to contain more information in orthography and in phonology and so are more complex to map than shorter words. Inconsistent words are harder to map because they benefit less than consistent words from learning mappings for other words with similar orthographic forms. Future investigation of the triangle model could determine the interplay between these factors and the extent to which they are explained by exposure, or require additional reconfiguring of architectural parameters.

The computational modelling approach demonstrated here enables isolation and control of various contributors to behavioural performance. In this respect it provides a useful accompaniment to approaches that demonstrate the observed correlations among various psycholinguistic variables. The computational modelling means that causal relations among these variables can be tested. In particular, we varied vocabulary size and exposure to measure how frequency effects vary between individuals. In the triangle model, exposure is the cause of variation in both vocabulary learning and frequency effects, in both first and second languages.

Acknowledgments

This research was supported by ESRC grant RES-000-22-4049.

References

Adelman, J. S., & Brown, G. D. A. (2008). Methods of testing and diagnosing model error: Dual and single route cascaded models of reading aloud. *Journal of Memory and Language, 59*, 524–544.

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word naming and lexical decision times. *Psychological Science, 17*, 814–823.

Adelman, J. S., Sabatos-DeVito, M. G., Marquis, S. J., & Estes, Z. (2014). Individual differences in reading aloud: A mega-study, item effects, and some models. *Cognitive Psychology, 68*, 113–160.

Andrews, S. (2015). Individual differences among skilled readers: The role of lexical quality. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 129–138).

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*(2), 283–316.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology, 58*, 412–424.

Brysbaert, M., Lagrou, E., & Stevens, M. (in press). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism: Language and Cognition* (in press).

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance, 42*, 441–458.

Chang, Y. N., Furber, S., & Welbourne, S. (2012). "Serial" effects in parallel models of reading. *Cognitive Psychology, 64*(4), 267–291.

Chateau, D., & Jared, D. (2000). Exposure to print and word recognition processes. *Memory & Cognition, 28*(1), 143–153.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256.

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review, 20*, 963–972.

Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2342 words. *Quarterly Journal of Experimental Psychology, 60*, 1072–1082.

Costa, A. (2005). Lexical access in bilingual production. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 308–328). New York: Oxford University Press.

de Groot, A. M. B., Borgwaldt, S., Bos, M., & van den Eijnden, E. (2002). Lexical decision and word naming in bilinguals: Language effects and task effects. *Journal of Memory and Language, 47*, 91–124.

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first and second language word recognition: A lexical entrenchment account. *Quarterly Journal of Experimental Psychology, 66*, 843–863.

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin & Review, 15*(4), 850–855.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition, 24*, 143–188.

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin, 125*(6), 777–799.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*, 627–635.

Garlock, V. M., Walley, A. C., & Metsala, J. L. (2001). Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and Language, 45*(3), 468–492.

Ghyselinck, M., Lewis, M. B., & Brysbaert, M. (2004). Age of acquisition and the cumulative-frequency hypothesis: A review of the literature and a new multi-task investigation. *Acta Psychologica, 115*, 43–67.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language, 58*(3), 787–814.

Gomez, P., & Perea, M. (2014). Decomposing encoding and decisional components in visual-word recognition: A diffusion model analysis. *The Quarterly Journal of Experimental Psychology, 67*(12), 2455–2466.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106*, 491–528.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meaning of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111*, 662–720.

Howes, D. H., & Solomon, R. L. (1951). Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology, 41*(6), 401–410.

Kaushanskaya, M., Yoo, J., & Marian, V. (2011). The effect of second-language experience on native-language processing. *Vigo International Journal of Applied Linguistics, 8*, 54–77.

Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *Quarterly Journal of Experimental Psychology, 68*(8), 1665–1692.

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research, 18*, 137–171.

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance, 39*(3), 802.

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwisterlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 12–31.

Lewellen, M. J., Goldinger, S. D., Pisoni, D. B., & Greene, B. G. (1993). Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision, and semantic categorization. *Journal of Experimental Psychology: General, 122*(3), 316–330.

Li, P., & Zhao, X. (2013). Self-organizing map models of language acquisition. *Frontiers in Psychology, 4*, 828.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science, 20*(12), 1507–1515.

Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics, 19*, 313–330.

Miller, G. A. (1990). WordNet: An on-line lexical database. *International Journal of Lexicography, 3*, 235–312.

Monaghan, P., & Ellis, A. W. (2010). Modeling reading development: Cumulative, incremental learning in a computational model of word naming. *Journal of Memory and Language, 63*, 506–525.

Morrison, C. M., Hirsh, K. W., Chappell, T., & Ellis, A. W. (2002). Age and age of acquisition: An evaluation of the cumulative frequency hypothesis. *European Journal of Cognitive Psychology, 14*(4), 435–459.

Norris, D. (2009). Putting it all together: A unified account of word recognition and reaction-time distributions. *Psychological Review, 116*, 207–219.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357–383.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review, 114*, 273–315.

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 539–557.

Pfost, M., Dörfler, T., & Artelt, C. (2013). Students' extracurricular reading behavior and the development of vocabulary and reading comprehension. *Learning and Individual Differences, 26*, 89–102.

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review, 107*(4), 786–823.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56–115.

Preston, K. A. (1935). The speed of word perception and its relation to reading ability. *Journal of General Psychology, 13*, 199–203.

Ransdell, S. E., & Fischler, I. (1987). Memory in a monolingual mode: When are bilinguals at a disadvantage? *Journal of Memory and Language, 26*, 392–405.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111*, 159–182.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*(3), 403–428.

Sears, C. R., Siakaluk, P. D., Chow, V. C., & Buchanan, L. (2008). Is there an effect of print exposure on the word frequency effect and the neighborhood size effect? *Journal of Psycholinguistic Research, 37*(4), 269–291.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523–568.

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *The Journal of Psychology, 9*(2), 371–377.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, 411–416.

Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language, 25*, 608–631.

Thorndike, E. L. (1931). *A teacher's word book of twenty thousand words*. New York: Teacher College, Columbia University.

Van Wijnendaele, I., & Brysbaert, M. (2002). Visual word recognition in bilinguals: Phonological priming from the second to the first language. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 616–627.

Wechsler, D. (2008). *Wechsler adult intelligence scale* (4th ed.). San Antonio, TX: Pearson.

Weide, R. L. (1998). The CMU pronouncing dictionary <http://www.speech.cs.cmu.edu/cgibin/cmudict>.

Welbourne, S. R., & Lambon Ralph, M. A. (2007). Using parallel distributed processing models to simulate phonological dyslexia: The key role of plasticity-related recovery. *Journal of Cognitive Neuroscience, 19*(7), 1125–1139.

Whitford, V., & Titone, D. (2012). Second-language experience modulates first-and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review, 19*(1), 73–80.

Yap, M. J., & Balota, D. A. (2009). Visual word recognition in multisyllabic words. *Journal of Memory and Language, 60*, 502–529.

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance, 38*(1), 53–79.

Yap, M. J., Balota, D. A., Tse, C. S., & Besner, D. (2008). On the additive effects of stimulus quality and word frequency in lexical decision: Evidence for opposing interactive influences revealed by RT distributional analyses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(3), 495–513.

Ziegler, J. C., Castel, C., Pech-Georgel, C., George, F., Alario, F. X., & Perry, C. (2008). Developmental dyslexia and the dual route model of reading: Simulating individual differences and subtypes. *Cognition, 107*, 151–178.